BINARY OPTIMIZATION:
APPLICATIONS TO REGIONAL PLANNING

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zevi Azzaino

August 2011

BINARY OPTIMIZATION:

APPLICATIONS TO REGIONAL PLANNING

Zevi Azzaino, Ph.D.

Cornell University 2011

This dissertation is based on three related essays applying binary optimization, with a focus on discovering the best selection of nodes, parcels, and villages that are subject to a budget constraint. The applications are special cases of a mathematical problem called the knapsack problem.

The first essay discusses the clustered knapsack problem with an underlying graph structure. Three different models are formulated in addition to the basic knapsack problem, and instance family groups of cluster types are constructed. A series of experiments solving the instances to find the optimal solutions are performed.

The second essay focuses on the study of rehabilitation and reconstruction of housing in Banda Aceh, Indonesia, which was destroyed by the tsunami and earthquake in 2004. Problems arose due to lack of planning, although millions of dollars were set aside for the reconstruction effort. New houses in the villages were badly constructed with minimal or no infrastructure support where they were located far from the closest rebuilt hospital, schools and other infrastructures. As such, one way to model this problem is to focus on individual villages to rebuild, building up their infrastructure so that everyone from that village and nearby villages can have access to the village community. The clustered knapsack models are applied to the City of Banda Aceh.

Finally, the third essay evaluates watershed protection in the Skaneateles Lake, which is the primary water supply for the city of Syracuse, NY. The high quality of

the water makes it possible to utilize the lake's water without filtration. The City of Syracuse was granted a filtration waiver by signing a Memorandum of Agreement, subject to several very strict conditions which include continuous monitoring of key water quality parameters, a back-up disinfection system, and a rigorous watershed protection program to reduce pathogen, chemical, nutrient and sediment loading into the lake. Part of the watershed management program involves the establishment of a riparian buffer at important areas within the watershed. We approach the riparian buffer problem using clustered knapsack models which will be applied to a selection of parcels in the Town of Skaneateles.

BIOGRAPHICAL SKETCH

Zevi Azzaino was born in Bogor, Indonesia on May 9, 1967. He is the second of three children of Farida and Zulkifli Azzaino. His mother was a kindergarten teacher and his father was a professor in Bogor Agricultural University (IPB). His father passed away when he was in high school, and his mother passed away recently in November 2009. He is married to Yuniati who gave him three sons, Zain (16), Zaki (11), and Zaidan Azzaino (9 months).

Zevi grew up in Bogor where he finished his formal education until high school, and went to Bandung for his undergraduate degree. In 1993, he received his Bachelor of Science in Environmental Engineering from Bandung Institute of Technology (ITB).

In 1994, he began his career as a government employee in the Central Project Management Office (CPMO) under the Directorate of Program Development within the Ministry of Public Works of Indonesia. As a Project Officer, he controlled the implementation of Integrated Urban Infrastructure Development Program (IUIDP) funded by Asian Development Bank (ADB) in the sectors of water supply, sanitation, urban roads and drainage; provided technical assistance to the local government, city and county offices; and assured compliance with ADB, the World Bank and the Government of Indonesia policies. Due to his achievements, he was awarded a scholarship funded by ADB to pursue a Master degree. In 2003, he received his M.Sc. in the field of Regional Science at Cornell University. He continued with his doctoral studies in the same field.

When the 2004 tsunami and earthquakes devastated Aceh and other parts of the world, together with his Cornell colleagues, Mazalan Kamis and Saiful Mahdi, he co-founded Aceh Relief Fund, a not-for-profit organization aimed to alleviate the sufferings of people in Aceh. The Fund raised nearly US$200,000. During his studies at Cornell, he served as guest lecturer and teaching assistant in several classes including Introduction to Geographic Information System (Professor Ann-Margaret Esnard), Urban Economics (Professor Kieran Donaghy), Analytical Mapping and Spatial Modeling (Professor Joe Francis), and Economics of Financial Crisis (Professor Iwan Azis). He has worked in the Intelligent Information System Institute (IISI) within the Department of Computer Science under the supervision of Professor Carla Gomes and Program of Applied Demographics (PAD) within the College of Human Ecology under the supervision of Professor Joe Francis. Zevi completed his Ph.D. in June 2011.

To

My wife Yuniati Zevi:

For her love, support, and faith

My sons Zain, Zaki and Zaidan Azzaino:

For the hope and promise of wonderful years ahead

and in loving memories of my parents

Zulkifli Azzaino and Farida Achmad:

For the encouragement, patience and as role models

ACKNOWLEDGMENTS

I would like to acknowledge the many individuals who provided assistance, guidance and support during my studies at Cornell.

First and foremost, I would like to express my sincere appreciation to Professor Jon Conrad, my special committee chairperson, for his guidance, precious comments, and constant encouragement during my studies. Jon's thoughtful considerations and understanding was instrumental in guiding me successfully through the Ph.D. program. My gratitude also goes to Professor Carla Gomes, who served on my committee, for her bright ideas, supervision and support. The opportunity to work with her, using advanced computing facilities in the Department of Computer Science, enhanced my research and accelerated the completion of my study. I would like to thank Professor Iwan Azis, who also served on my committee, for his contributions and mentorship. Iwan has been an outstanding advisor and a good friend, constantly attentive in my academic life at Cornell. I feel very fortunate to have worked with such a wonderful committee.

I appreciate Professor Kieran Donaghy, Director of Graduate Studies of Regional Science, for his valuable counsel and financial support through assistantship at Cornell. I also wish to thank Professor Ann-Margaret Esnard, who introduced me to GIS and provided many great opportunities as guest lecturers in her class during my first years of my program. I would also like to thank Professor Joe Francis for his support, friendship, and opportunity.

Many thanks to the staff in the Department of City and Regional Planning: Donna Wiernicki, Tina Nelson, Lorene Walker and Chaterine Lopez, for their support.

Special appreciation to Mr. Djoko Kirmanto, Ministry of Public Works, who gave permission and approved the scholarship for my Master and Ph.D. program at

Cornell. I would like to thank to all my colleagues at the Ministry of Public Works and the Ministry of Public Housing for their friendship and support.

I would like to thank to my fellow Ph.D. students, Katherine Lai for her assistance in Cplex programming, and Inka Yusgiantoro for his help in collecting data and information for the Banda Aceh essay. I am grateful to Lee MacBeth, City of Syracuse's Watershed Control Coordinator, for her support in providing data and valuable information for the Skaneateles Lake essay.

Most importantly, my never-ending gratitude goes to my wonderful wife, Yuniati for her loving support and patience. Her willingness to come to Cornell, and pursuing her own Ph.D. degree while raising our three children: Zain, Zaki, and Zaidan, was an amazing gift which I will always be grateful. Also, I would like to remember my mother, who passed away in November 2009, my late father, and my sister, who all were very supportive and always prayed for my success.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

Introduction and Overview

This dissertation is based on three related essays applying binary optimization, with a focus on discovering the best selection of nodes, parcels, and villages that are subject to a budget constraint. The applications are special cases of a mathematical problem called the knapsack problem. The name derives from the problem of choosing items to fit into a knapsack, when the items have different potential utility and cost where the carrier has a weight constraint on how much he or she can carry.

The first essay discusses the *clustered knapsack problem*, which is a generalization of the knapsack problem with an underlying graph structure. Given a graph $G = (V,E)$, there are nonnegative costs c_v and values (or utilities) u_v for each node $v \in V$. In addition, there is a set of clusters $C_k \subset V$, $k = 1, \ldots, m$. Each cluster C_k also has some utility u_k which is included if and only if all nodes in the cluster have been included in the knapsack of size M. Three different models are formulated in addition to the basic knapsack problem. We construct instance families with 100 datasets, each containing 100 nodes. We perform a series of experiments solving the instances to find the optimal solutions. This work will be used as a basis for applications in the next two essays.

The second essay focuses on the study of rehabilitation and reconstruction of housing in Banda Aceh, Indonesia, which was destroyed by the tsunami and earthquake in 2004 which wiped out the villages and displaced more than 100,000 people. Problems arose due to lack of planning, although millions of dollars were set aside for the reconstruction effort. Because not every infrastructure needed was rebuilt at once, thousands of people remained homeless even after two years had passed. New houses in the villages were badly constructed, lack running water, had no roads to town, and were still surrounded by the rubble of neighboring houses not yet rebuilt, or were far from the closest rebuilt hospital, schools and other infrastructures. As such, one way to model this problem is to focus on individual villages to rebuild, building up their infrastructure so that everyone from that village can go home and have access to the village community. There is also an underlying benefit for clustering the rebuilt villages such that villages next to a rebuilt village with a hospital can now have access to this hospital and potentially other resources such as schools and other infrastructure as well. The clustered knapsack models will be applied to 89 villages in the City of Banda Aceh.

Finally, the third essay will evaluate watershed protection in the Skaneateles Lake using the clustered knapsack models developed in the previous essay. Skaneateles Lake is the primary water supply for the city of Syracuse, NY. It is one of the cleanest lakes in the Finger Lakes, and is located approximately 20 miles southwest of the city. The high quality of the water makes it possible to utilize the lake's water without filtration. The City of Syracuse was granted a filtration waiver by signing a Memorandum of Agreement, subject to several very strict conditions which include continuous monitoring of key water quality parameters, a back-up disinfection system, and a rigorous watershed protection program to reduce pathogen, chemical, nutrient and sediment loading into the lake. Part of the watershed management program involves the establishment of a *riparian buffer* at important areas within the watershed. We will approach the riparian buffer problem using clustered knapsack models, which will be applied to a selection of parcels in the Town of Skaneateles with a total of 1834 parcels covering 12,340 acres and 52 different land-uses.

CHAPTER TWO

Project to Study Clustered Knapsack

2.1 Introduction

The knapsack problem is a classical NP-complete problem and is one of Karp's 21 NP-complete problems (Karp, 1972). Given a set of items indexed 1, …, v, …, n, each with a utility value u_v and size s_v, and a knapsack size S, the goal is to find a subset W of the items such that $\sum_{v \in W} s_v \leq S$ and the quantity $\sum_{v \in W} u_v$ is maximized. While the problem is NP-hard, it admits a dynamic programming algorithm (Lawler, 1979). When a dynamic programming algorithm exists, the problem is weakly NP-hard, meaning that it can be solved in polynomial time if the size of the knapsack S is assumed to be constant.

In this work, we introduce the clustered knapsack problem, a generalization of the knapsack problem with an underlying graph structure. Given a graph $G = (V,E)$, there are nonnegative costs c_v and values (or utilities) u_v for each node $v \in V$. In addition, there is a set of clusters $C_k \subset V$, $k = 1, \ldots, m$. Each cluster C_k also has some utility u_k which is included if and only if all nodes in the cluster have been included in the knapsack of size M. The goal is then to find a subset W of the nodes that satisfies $\sum_{v \in W} c_v \leq M$ and maximizes the sum of the utilities of the included nodes and cluster $\sum_{v \in W} u_v + \sum_{C_k \subset W} u_k$.

This problem is motivated by applications in the optimization of the riparian buffer in the Skaneateles Lake watershed, and the reconstruction of villages in Banda Aceh of Indonesia after the tsunami of 2004. The City of Syracuse uses the high quality of Skaneateles Lake water as the source for drinking water. The water is of such high quality that the City does not need to build a filtration plant that would cost more than $70 million. To protect the water quality and to satisfy the provisions of the

EPA, the City uses riparian land buffers at critical areas as an important conservation tool. The selection of land or parcels to be included in the riparian buffer is based on parcel's score or ranking system developed by the City of Syracuse (Azzaino et al., 2002). These parcels are often farther away from the intake, which is one of the key factors in the ranking system. Furthermore, these selected parcels' locations are far apart. When a parcel is acquired, its neighboring parcels may still runoff pollutants through the acquired parcel to the stream or lake. Thus there is an essential benefit for clustering the acquired parcels to avoid such problems.

In the case of Banda Aceh, the tsunami disaster wiped out the villages and displaced more than 100,000 people. Millions of dollars were set aside for the reconstruction effort, but because not everything was (nor could be) rebuilt at once, problems arose due to lack of careful planning. In particular, many thousands of people remained homeless even after two years had passed. While some homes were rebuilt, the new houses were oftentimes useless because they were shoddily constructed, still had no running water, had no roads to town, were still surrounded by the rubble of neighboring houses not yet rebuilt, or were far from the closest rebuilt hospital, schools and other infrastructure. As such, one way to model this problem is to focus on individual villages to rebuild, building up their infrastructure so that everyone from that village can go home and have access to the village community. There is also an underlying benefit for clustering the rebuilt villages such that villages next to a rebuilt village with a hospital can now have access to this hospital and potentially share other resources, such as schools and markets as well. Not all villages can be rebuilt at once due to limited resources, and so the underlying computational problem is that of selecting which villages to rebuild first.

A different form of this knapsack problem has been previously studied as the Partially Ordered Knapsack (POK) problem, where there is a partial order on the universe of items, and each item can only be put in the knapsack if all of its ancestors are also in the knapsack. Johnson et al. (1983) showed that there exists a dynamic programming algorithm when the dependency graph of the partial order is a tree and that it is strongly NP-hard when it is not.

2.2 Problem Definition

We will first define the clustered knapsack problem as follows:

Input:

- S, a set of n items:

$$S = \{s_v \mid v = 1, 2, \ldots, n\};$$

  Each item s_v is characterized by a utility u_v and a cost c_v.

- C, a set of m clusters, each cluster is a subset of items:

$$C = \{C_k : C_k \subseteq S \mid k = 1, 2, \ldots, m\};$$

  Each cluster C_k is characterized by a utility u_k.

- M, a constant representing the capacity of the knapsack

Output:

- A set of items $Z^* \subseteq S$
- A set of clusters $Q^* \subseteq C$

Such that:

– $\forall s_k \in Z^*, \forall C_q \in Q^*;$

– $$(Z^*, Q^*) = argmax(Z,Q)\, f(Z,Q)$$
$$= arg\,max(Z,Q) \sum_{s_i \in Z} u_i + \sum_{C_q \in Q} u_k \;\; s.t. \sum_{s_i \in Z} c_i \le M$$

2.3 Problem Formulation

We can formulate the problem defined in 2.2 as a Mixed Integer Program (MIP). A mixed-integer program is the minimization or maximization of a linear function subject to linear constraints on a set of variables, where some of the variables may be constrained to take on only integer values (Nemhauser, 1979). The problem formulation for our problem is as follows:

$$\text{Maximize } \sum_{v} u_v x_v + \sum_{k} u_k y_k \tag{1}$$

Subject to:

$$\sum_{v} c_v x_v \leq M \tag{2}$$

$$y_k \leq x_v;\ \forall k,\ v \in C_k \tag{3}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{4}$$

$$\mathrm{y}_k \geq 0;\ \forall C_k \tag{5}$$

$$v = 1, 2, \ldots, n;\ \mathrm{k} = 1, 2, \ldots, m;$$

Where:

x_v: an integer variable indicating whether or not node v is bought (1 if node v is bought, 0 otherwise)

y_k: a real variable indicating whether or not all members of the cluster C_k have been bought

The problem parameters of this model are the following:

- Items:
  - n – number of items
  - for each item s_v:
    - u_v – utility of item s_v;
    - c_v – cost of item v;

- Clusters
    - m – number of clusters;
    - for each clusters C_k:
        - Set of items that constitute the cluster C_k;
        - u_k – utility of cluster C_k;
- Knapsack
    - M – capacity of knapsack;

2.4 Synthetic Instances

In order to generate synthetic instance sets for studying the Clustered Knapsack problem we generate random values for the different problem parameters. We consider a model in which the utilities and cost of the items are weakly correlated. In this model, in addition to the problem parameters listed above we will consider:

- l – amplitude of the interval for the item costs; the cost of each item, (c_v) is generated uniformly at random [1 l]
- d – amplitude of the interval for the item utilities, w.r.t. the corresponding item cost; the utility of each item (u_v) is generated uniformly at random from [c_v – d c_v + d];
- a – ($0 \le a \le 1$) proportion of the total cost of the items that defines M;

$$M = a \sum_{s_v \in S} c_v \qquad (6)$$

- a graph G = (S, E), in which the node set S corresponds to the set of items, and the edge set E denotes adjacency relationships between items. The clusters are defined based on this graph.

- ε – (ε ≥ 0) weight of the total utility of the cluster items that defines the utility of the cluster. In other words, ε is a parameter to quantify the benefit of utility of producing or having a cluster.
- N – number of problem instances in a set;

An instance of the Clustered Knapsack problem is therefore defined by a tuple:

$$\langle \text{n, l, d, a, } \varepsilon\text{, E, C} \rangle$$

Where:

- n – number of items
- *l* – amplitude of the interval for the item costs; the cost of each item, (*c*) is generated uniformly at random [1 *l*]
- d – amplitude of the interval for the item utilities, w.r.t. the corresponding item cost; the utility of each item (*u*) is generated uniformly at random from [c_j – d c_j + d];
- a – ($0 \leq a \leq 1$) proportion of the total cost of the items that defines M; $M = a\sum_{s_j \in S} c_j$
- a graph G = (S, E), in which the node set S corresponds to the set of items, and the edge set E denotes adjacency relationships between items.
- C – set of items that define each cluster; this will be based the graph G = (S, E);
- ε – (ε ≥ 0) weight of the total utility of the cluster items that defines the utility of the cluster.
- N – number of problem instances in a set;

Note that the parameters can be changed to generate different instances. However, we will keep most of the parameters fixed (at least initially), focusing on:

- Clusters types:
    - **Neighbor** – in this type each item (node in the graph) induces a cluster; the cluster is made out of: (1) the node that induces the cluster and (2) all the nodes adjacent to it.
    - **Edge** – in this case a cluster corresponds to a pair of adjacent items or nodes. Therefore all the clusters have size two. We can see this case as having connectors between adjacent items (or nodes or plots) and if both of the items (or plots or nodes) are selected the connector will be on. This may be a more natural way of modeling the notion of adjacency.
- Sensitivity analysis of the clustering effect with respect to the utility assigned to each cluster. To analyze this aspect, for each family of instances we will consider different values for the parameter ε, and solve the instances using IBM/Ilog's Cplex solver (ILOG, 2007).

2.4.1 Instance Families – Neighbor Type

$$\langle n, l, d, a, \varepsilon, E, C \rangle$$

Where:

- n = 100
- l = 1000
- d = 100
- a = 0.3
- graph G = (V, E) – a square lattice, 10x10. Each cell in this lattice corresponds to an item, a vertex in V; for every pair of adjacent cells there is an edge in E.

- C – each vertex *v* in V induces a cluster; the cluster induced by vertex *v* includes the vertex *v* and all other vertices in V that share an edge with *v*. Therefore there are 100 clusters; 64 clusters, each made out of 9 vertices; 4 clusters (corresponding to the corners of the lattice), each made out of 4 vertices; 32 clusters (corresponding to the borders of the lattice, excluding the corners), each made out of 6 vertices.
- N = 100
- ε = 0.00 (initial value);

Each family of instances is characterized by a given ε, and all the other parameters are fixed; in order to generate different families of instances in a given type, we vary the value of ε; basically we identify the value of ε that causes an abrupt change in terms of the clustering effects – i.e., that causes the solution to go from mainly unclustered to highly clustered. Therefore in this type we will generate (100 × Number-of-ε-values) instances.

2.4.2 Instance Families – Edge Type

$$\langle n, l, d, a, \varepsilon, E, C\rangle$$

Where:

- n = 100
- l = 1000
- d = 100
- a = 0.3
- graph G = (V, E) – a square lattice, 10x10. Each cell in this lattice corresponds to an item, a vertex in V; for every pair of adjacent cells there is an edge in E.

- C – each vertex *v* in V induces as many clusters as the number of edges involving node *v*; each cluster, induces by a vertex *v*, includes two nodes: *v* and a node connected to it; therefore given a 10 × 10 lattice for example, there are 64 × 8 + 32 × 5 + 4 × 3 = 684 clusters, all of size 2.
- N = 100
- ε = 0.00 (initial value);

Again, each family of instances is characterized by a given ε, and all the other parameters are fixed; in order to generate different families of instances in a given type, we vary the value of ε; and we identify the value of ε that causes an abrupt change in terms of the clustering effects – i.e., that causes the solution to go from mainly unclustered to highly clustered. Therefore, as in Neighbor type, in this type we will generate (100 × Number-of-ε-values) instances.

2.5 The Models

2.5.1 The Basic Knapsack Model

The weight of the total utility of the cluster items, ε, defines the utility of the cluster, $\forall C_k \in C$

$$u_k = \varepsilon \sum_{v \in C_k} u_v \tag{7}$$

When ε = 0 (initial value), we do not consider any cluster benefit and therefore Equation (1) becomes

$$\text{Maximize} \sum_v u_v x_v \tag{8}$$

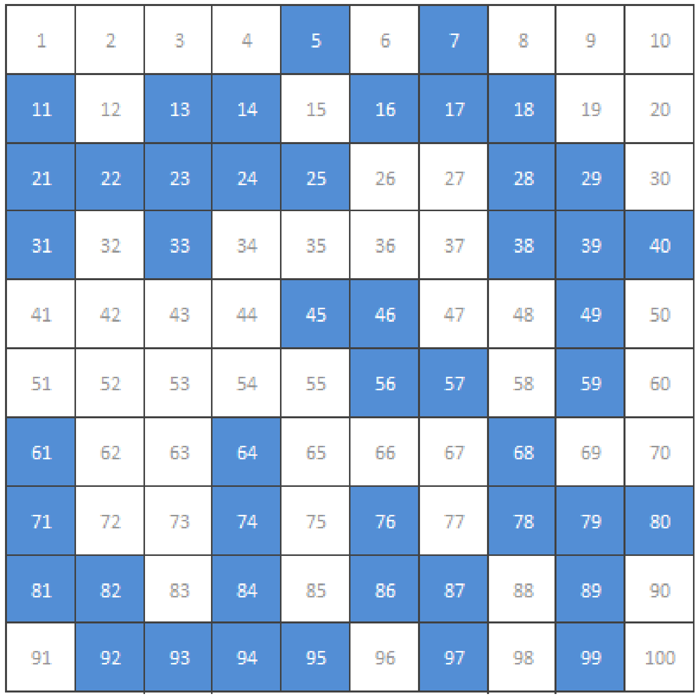
$$\text{s.t.} \quad \sum_v c_v x_v \le M \tag{9}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{10}$$

We refer to Equation (8) as the **Basic Knapsack Model**. We perform N=100 series of experiments with a 10x10 grid as stated previously and solve the instances to optimality using Cplex. Run time to find the optimal solution for all 100 Ns (instances) are between 0 to 0.02 seconds with 13 instances having a runtime of 0 seconds, 80 instances 0.01 seconds and 7 instances 0.02 seconds.

Figure 2.1 shows the visualization of optimal solution for two instances for the Basic Knapsack problem, instance N=9 and 57. The blue nodes were the selected nodes, while white were not selected.

A.



B.

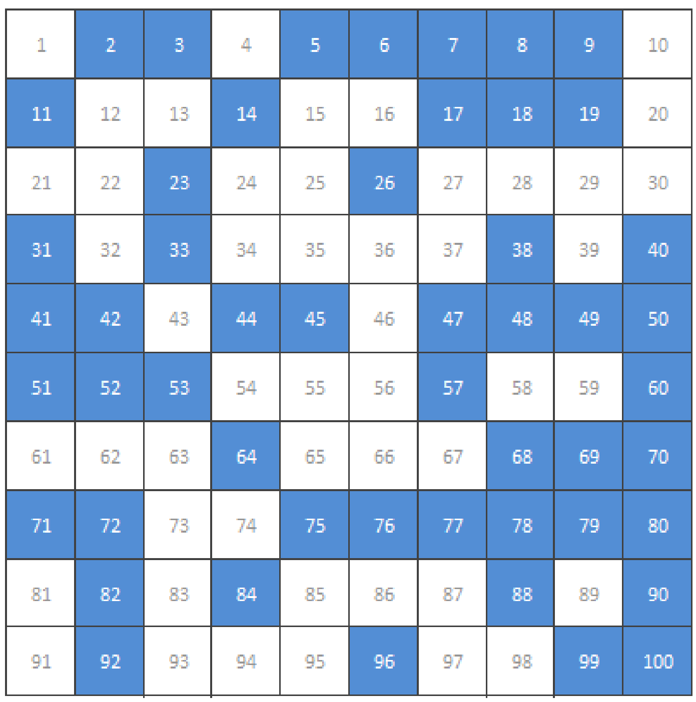


Figure 2.1. Visualization of the Optimal Solution for the Basic Knapsack problem

A. N= 9, total Utility= 31139, total cost= 15801, run time= 0.02 seconds, nodes selected= 47

B. N= 57, total Utility= 34882, total cost= 13762, run time= 0.01 seconds, nodes selected = 51

2.5.2 The Basic Clustered Knapsack Model

Eliminating the first part of Equation (1) we get only the cluster utility benefit as follows:

$$\text{Maximize } \sum_k u_k y_k \tag{11}$$

$$\text{s.t.} \quad \sum_v c_v x_v \leq M \tag{12}$$

$$y_k \leq x_v;\ \forall k,\ v \in C_k \tag{13}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{14}$$

$$y_k \geq 0;\ \forall C_k \tag{15}$$

Where

$$u_k = \sum_{v \in C_k} u_v \tag{16}$$

We refer to Equation (11) as the **Basic Clustered Knapsack Model**. We perform N=100 series of experiments for each family types: Neighbor and Edge, and solve the instances to optimality using Cplex. The optimal solutions of the model running the two instance family types: Neighbor and Edge, are visualized in Figure 2.2 and 2.3.

Within the same instance N, the total utility of type Edge is always greater than type Neighbor. For example, the first instance of type Edge has a total utility of 91370, whereas the total utility for the type Neighbor has only a total utility of 77355. On the other hand, the running time for the type Edge is always faster than the type Neighbor, as shown in Table 2.1.

Table 2.1. Summary Results of the Basic Clustered Knapsack for All 100 Instances

	Total Utility		Running Time		Total Cost		# Nodes Bought	
	Edge	Neighbor	Edge	Neighbor	Edge	Neighbor	Edge	Neighbor
Min	85068	72934	0.01	0.03	13438	13414	38	34
Max	115908	107442	0.45	0.94	17692	17689	50	45
Ave	99551.86	87261.32	0.10	0.53	15675.77	15672.04	44.65	38.57
Median	100382.50	87410.50	0.09	0.57	15759.00	15750.50	45	39
StDev	6294.15	6549.23	0.07	0.21	944.59	942.44	2.41	2.57

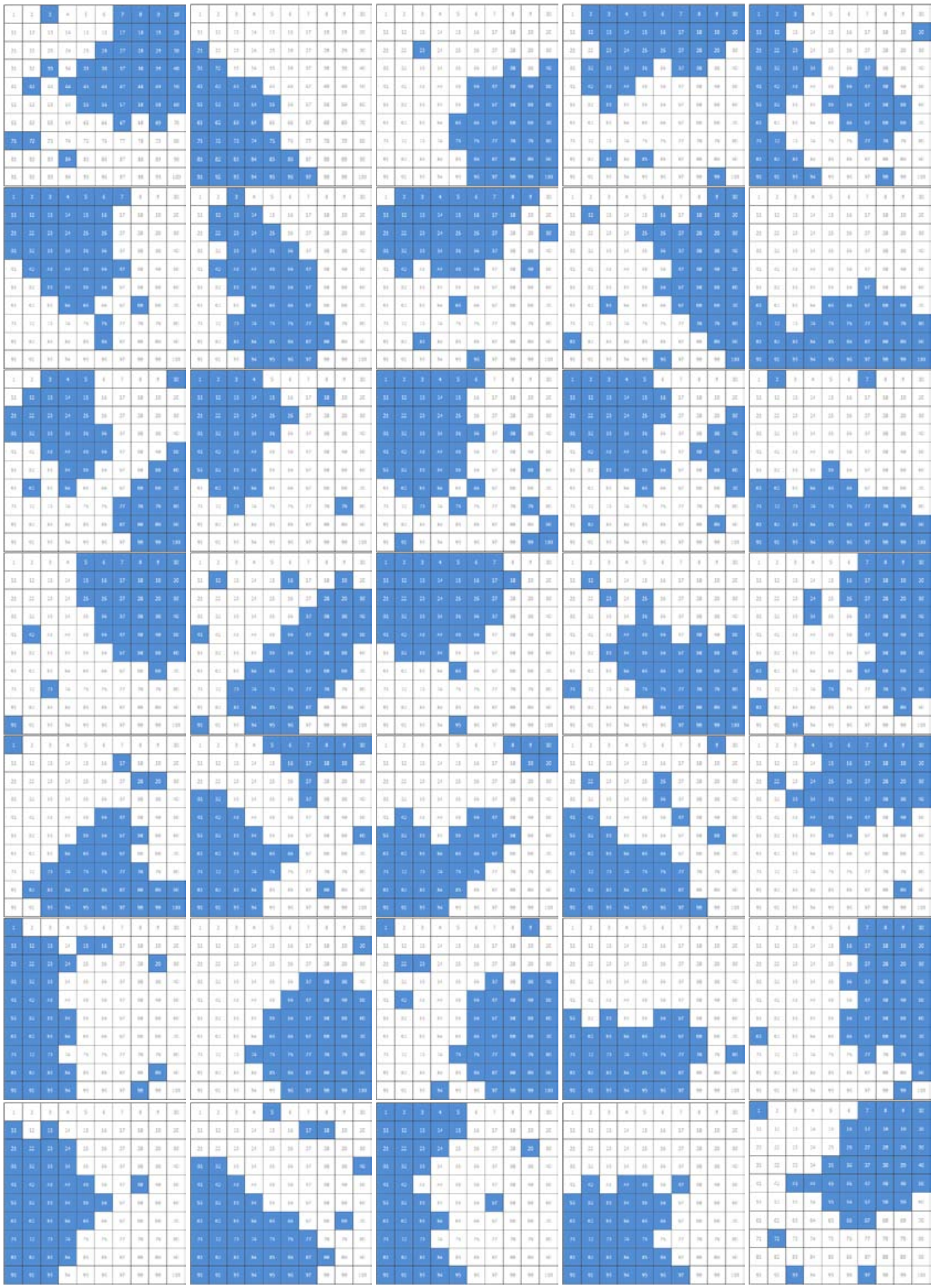


Figure 2.2. Visualization of the Optimal Solution for the Basic Clustered Knapsack, Cluster Type: Neighbor (Instances 0 through 34)

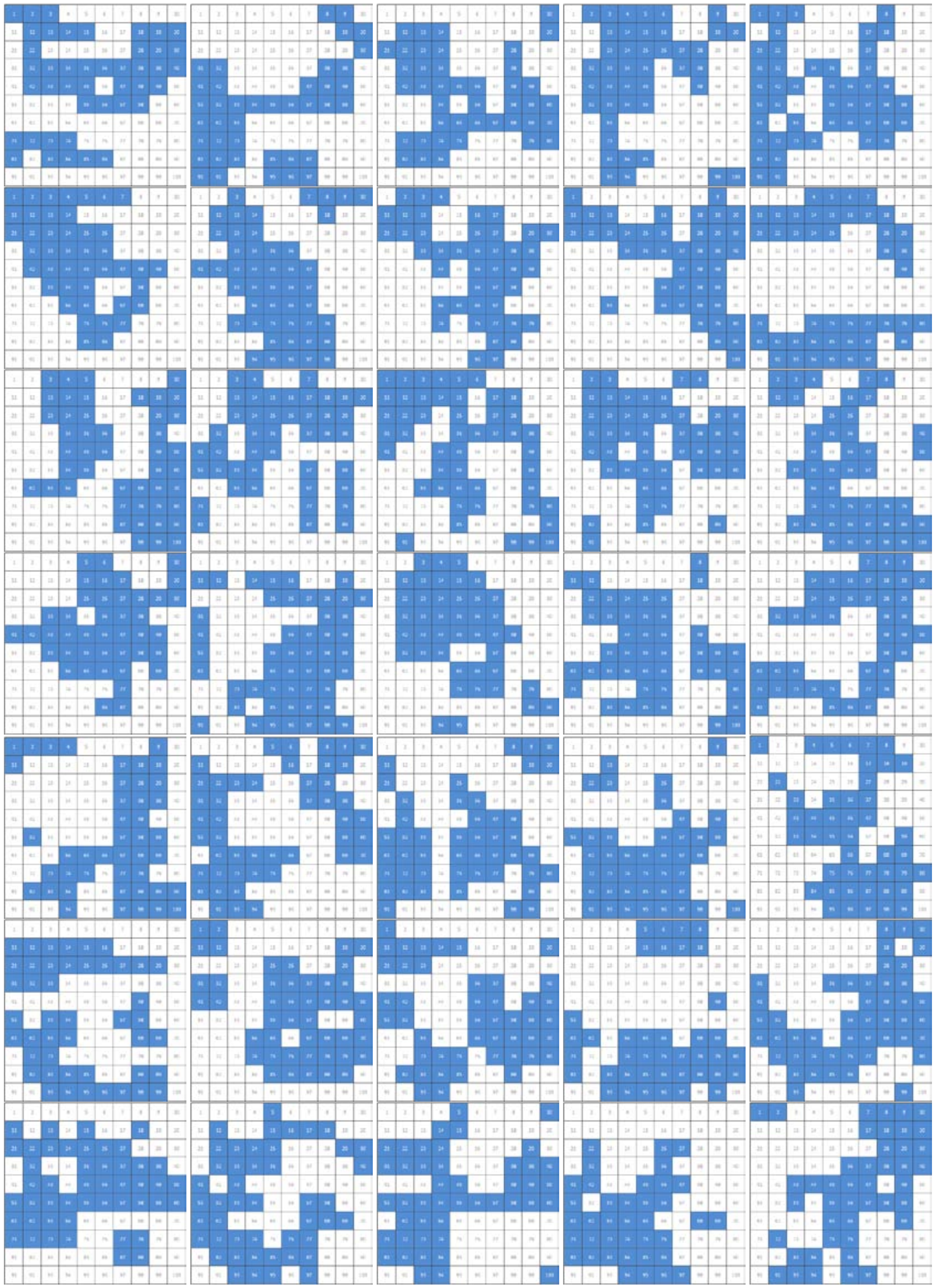


Figure 2.3. Visualization of the Optimal Solution for the Basic Clustered Knapsack, Cluster Type: Edge (Instances 0 through 34)

2.5.3 The Clustered Knapsack Model

Equation (1) is restated below for clarity.

$$\text{Maximize} \sum_{v} u_v x_v + \sum_{k} u_k y_k \tag{17}$$

$$\text{s.t.} \quad \sum_{v} c_v x_v \le M \tag{18}$$

$$y_k \le x_v;\ \forall k,\ v \in C_k \tag{19}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{20}$$

$$\mathrm{y}_k \ge 0;\ \forall C_k \tag{21}$$

The weight of the total utility of the cluster items, ε, defines the utility of the cluster, $\forall C_k \in C$

$$u_k = \varepsilon \sum_{v \in C_k} u_v \tag{22}$$

We refer to Equation (17) as the **Clustered Knapsack Model**. When ε > 0, double counting utility u_v occurs. In each family type we solve the instances to optimality using Cplex by varying the value of parameter ε starting from 0.01 up to 100 in steps 0.01, and identify the value of ε that causes a change in the clustering effects. For example, the first instances of type Neighbor (N=0), the values of ε that cause a change in the optimal solution are 0.01, 0.02, 0.04, 0.06, 0.09, 0.12, 0.15, 0.24, 0.26, 0.49, 0.68 and 2. This means that when ε = 0.06, the optimal solution does not change until ε = 0.09. Similarly, when ε = 2 the optimal solution does not change even at ε equals to 100. Table 2.2 shows the results of Clustered Knapsack Model in which the value of parameter ε causes a change in the optimal solution.

Table 2.2. Summary Results of the Clustered Knapsack based on Parameter ε that Changes the Optimal Solutions

	Parameter ε		Total Utility		Running Time		Total Cost		# Nodes Bought	
	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.
Min	0.01	0.01	27729.6	27324.4	0.01	0.01	13396	13388	33	30
Max	100	96	6340720	6832510	0.75	3.11	17710	17700	54	54
Average	1.09	2.24	105797.50	175935.86	0.09	0.27	15684	15658	45.69	43.30
StDev	5.03	10.92	329790.03	736623.88	0.06	0.37	930.54	933.92	2.70	4.13

The highest running time for the Clustered Knapsack problem in the type Neighbor is 3.11 seconds, and was obtained in N=37 at ε = 87. For instance N=37, the values of parameter ε that cause a change in the optimal solutions are ε = 0.01, 0.06, 0.07, 0.08, 0.11, 0.16, 0.17, 0.22, 0.28, 0.56, 0.64, 0.73, 86, and 87. The visualization of the optimal solution for Clustered Knapsack Model with type Neighbor at instance N=37, is shown in Figure 2.4.

It is interesting to note that up until ε= 0.64, most selected nodes clustered in the lower right end of the 10x10 grid, however when ε= 0.73 the selected nodes move to the upper left of the grid and remain there until ε= 87.

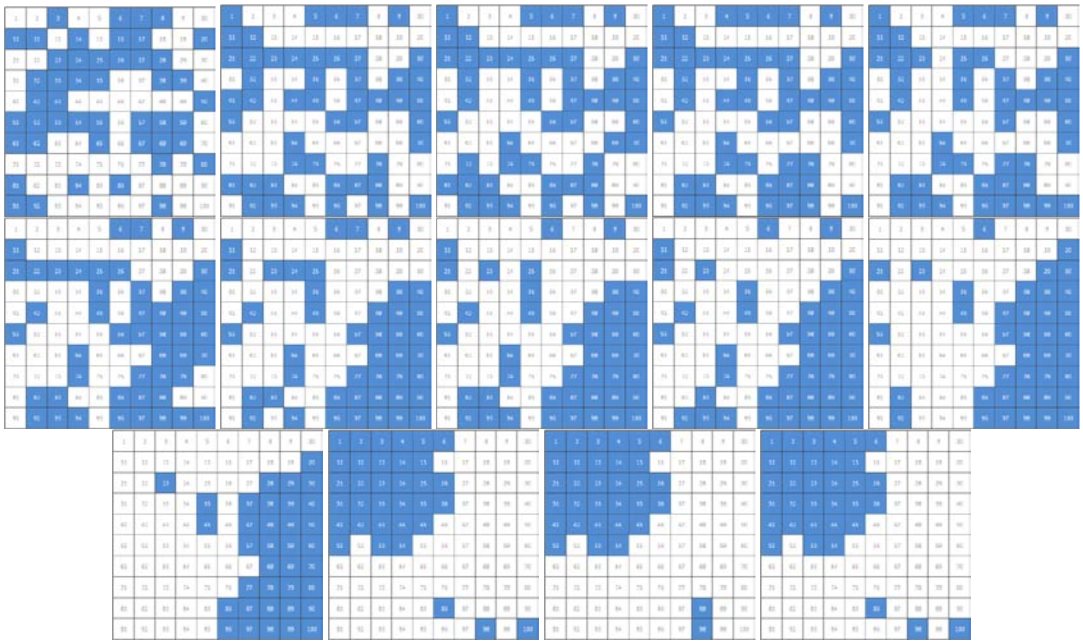


Figure 2.4. Visualization of the Optimal Solution for the Clustered Knapsack, Cluster Type: Neighbor (Instance, N= 37)

For type Edge, the highest running time is 0.75 seconds, which was obtained in instance N= 90 where ε = 1.9 (which is not the highest value of ε). For N= 90, the values of parameter ε that cause a change in the optimal solutions are ε = 0.01, 0.02, 0.09, 0.22, 0.24, 0.30, 0.49, 1.90 and 4.50. Figure 2.5 visualizes the optimal solution for the Clustered Knapsack Model with type Edge (N= 90).

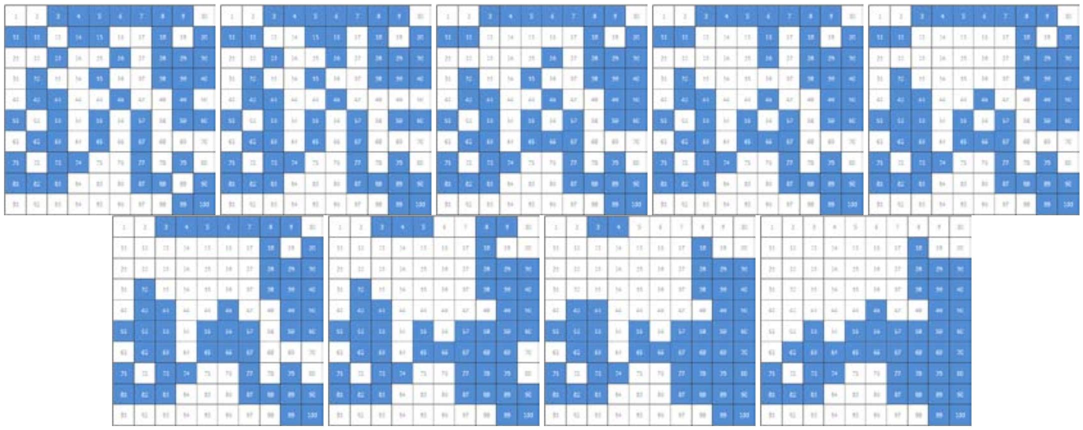


Figure 2.5. Visualization of the Optimal Solution for the Clustered Knapsack, Cluster Type: Edge (Instance, N= 90)

2.5.4 The Modified Clustered Knapsack Model

The objective function of the Clustered Knapsack Model is restated below:

$$\text{Maximize } \sum_v u_v x_v + \sum_k u_k y_k \tag{23}$$

Note that there is a possibility of double counting utility of the objective function. To avoid double counting of the utility, we replace the second term of Equation (23) with the following equation:

$$\sum_k \varepsilon \prod_{v \in C_k} x_v \tag{24}$$

Parameter ε is a weight to quantify the benefit of utility of producing a cluster. When all of the items in cluster k are selected, the product of x_v equals to 1, on the other hand if one or more items in cluster k are not selected, then the product of x_v equals to 0. Normally, using a product of variables in the MIP would cause the program to become nonlinear. However, we can still use the variables y_k introduced previously to indicate whether all nodes in cluster k have been selected when under the assumption that we have found an optimal solution that would push each value of y_k to the correct value.

Thus Equation (24) is replaced with the following equation:

$$\sum_k \varepsilon y_k \tag{25}$$

The objective function in Equation (23) therefore becomes:

$$\text{Maximize } \sum_v u_v x_v + \sum_k \varepsilon y_k \tag{26}$$

$$\text{s.t.} \quad \sum_v c_v x_v \leq M \tag{27}$$

$$y_k \leq x_v;\ \forall k,\ v \in C_k \tag{28}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{29}$$

$$\mathrm{y}_k \geq 0;\ \forall C_k \tag{30}$$

We refer to Equation (26) as the **Modified Clustered Knapsack Model**. We run the model for each family type: Neighbor and Edge with Cplex by varying ε from 0.01 to 1 with steps 0.01, continued with ε from 1.1 to 10 with steps 0.1, and ε from 11 to 100 with steps 1 and finally ε from 111 to 1000 with steps 10. We identify the value of ε that causes a change in the optimal solutions. For example, the eleventh instance in type Edge (N=10), the values of ε that cause a change in the optimal solutions are 0.01 (initial value), 16, 66, 99, 130, 160, 220, 480, 490, and 750. The results are shown in Table 2.3.

Table 2.3. Summary of Results of Modified Clustered Knapsack Based on Parameter ε that Changes the Optimal Solution

	Epsilon		Total Utility		Running Time		Total Cost		# Nodes Bought	
	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.	Edge	Neigh.
Min	0.01	0.01	27303.3	27303	0.01	0.01	13379	13388	40	37
Max	1000	1000	91197	51154	0.80	2.74	17706	17707	55	53
Average	238.08	294.32	43414.85	34647.72	0.08	0.18	15672.93	15668.30	47.70	45.58
StDev	243.66	259.56	13676.56	4563.72	0.07	0.23	928.55	927.79	2.40	2.57

The highest running time for the Modified Clustered Knapsack problem in type Neighbor is 2.74 seconds, which was obtained in N=20 where ε = 960. The values of parameter ε that cause a change in the optimal solutions are ε = 0.01, 0.21, 26, 56, 75, 310, 380, 600, 630, 960, and 1000. Figure 2.6 visualizes the optimal solution for the Modified Clustered Knapsack Model with cluster type Neighbor (N=20). Similarly, the type Edge for the Modified Clustered Knapsack Model is shown in Figure 2.7.

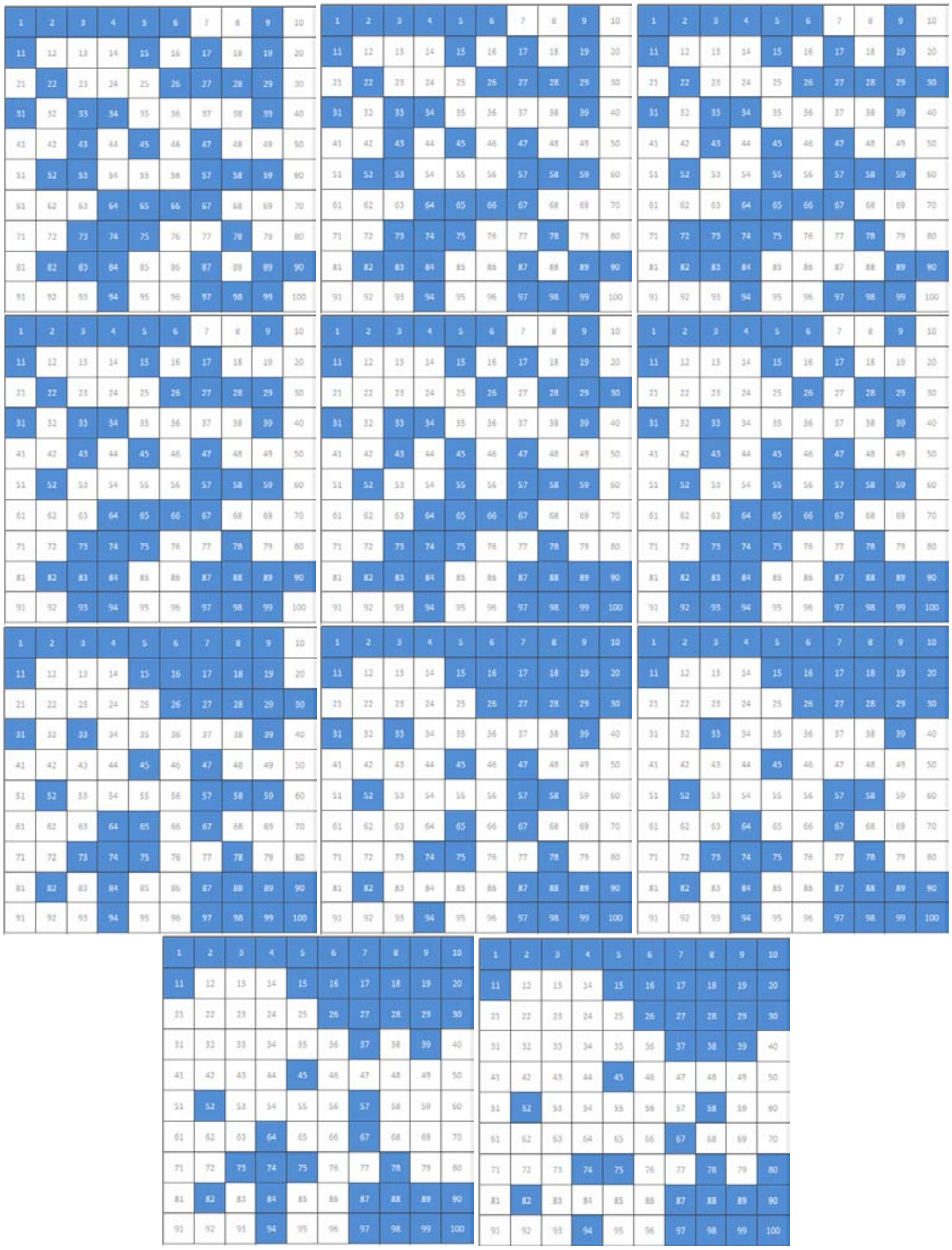


Figure 2.6. Visualization of the Optimal Solution for the Modified Clustered Knapsack, Cluster type: Neighbor (Instance, N= 20)

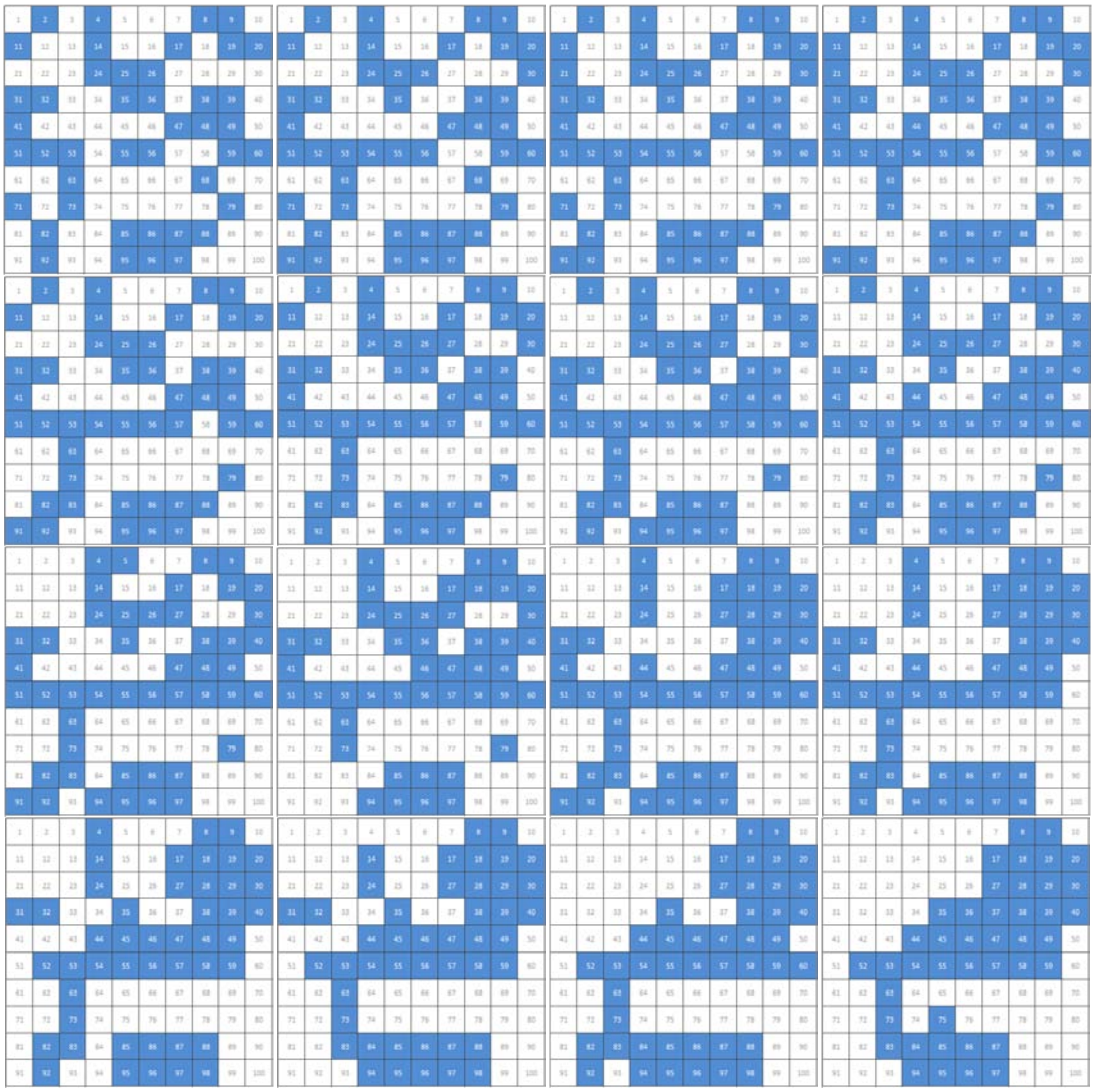


Figure 2.7. Visualization of the Optimal Solution for the Modified Clustered Knapsack, Cluster type: Edge (Instance, N= 1)

2.6 Summary

In this work, we investigate different mathematical formulations of the budget-constrained variant of the Clustered Knapsack Problem. In addition to the basic knapsack problem we proposed three other approaches which we refer to as Basic Clustered Knapsack, Clustered Knapsack and Modified Clustered Knapsack problem. We evaluate the performance of the models on 100 instances of 10x10 random grids and find the optimal solutions for different budgets and values of parameter ε. The results confirm that a higher parameter ε causes the solution to go from mainly unclustered to highly clustered nodes, on both neighbor and edge types.

This general optimization problem can also be used to model other applications in areas such as social networks and wildlife conservation. Finally, this work will be use as a basis of applications in the following chapters, the optimization of riparian buffer in the Skaneateles Lake watershed and the reconstruction of villages in Banda Aceh of Indonesia after the tsunami of 2004.

REFERENCES

Azzaino, Z., Conrad, J.M., Ferarro, P., 2002. Optimizing the Riparian Buffer: Harold Brook in the Skaneateles Lake Watershed, New York. Land Economics, 78(4): 501-514.

ILOG, SA. 2007. CPLEX 11.0 Reference Manual

Karp, R.M., 1972. Reducibility among combinatorial problems, in Complexity of Computer Computations. R. E. Miller and J. W. Thatcher, Eds. New York: Plenum, pp. 85-104.

Lawler, E.L., 1979. Fast approximation algorithms for knapsack problems. Mathematics of Operations Research, 4:339-356.

Nemhauser, G.L., Wolsey, L.A., 1988. Integer and Combinatorial Optimization, John Wiley, New York.

CHAPTER THREE

No Virtue of Necessity: Post-Tsunami Housing Reconstruction in Aceh

3.1 Introduction

When the tsunami crashed into the coastal communities of Aceh, Indonesia, in late 2004, some 227,898 people were killed, including those missing and presumed dead, more than 425,000 people were rendered homeless, and thousands of children were orphaned. The tsunami was triggered by an earthquake with a magnitude of between 9.1 and 9.3, the second largest earthquake ever recorded on a seismograph (Lay, 2005; Geist, 2005), and lasted for 8 to 10 minutes, the longest duration of faulting ever observed (Walton, 2005.) Countries hard hit by the disaster were Indonesia, Sri Lanka, India, and Thailand. However, Indonesia was by far the worst country hit. The costs of reconstruction were estimated at $4-5 billion for the period of five years, although it is unclear whether such estimates cover only the reconstruction of physical assets such as housing, roads, and infrastructure destroyed by the tsunami, or whether they also include the replacement costs for the non-physical losses.

Economically speaking, the supply-side loss can be offset by the demand-side expenditure spurred by the increased reconstruction spending. The latter is funded largely by the financial assistance from abroad. In early 2005, not long after the disaster took place, the United Nations secretary general called for $977 million in immediate funds for the next 6 months' emergency relief for tsunami victims. Pledges amounting to US$2 billion came in from 44 countries, the biggest from Japan, with US$500 million (*Financial Times* 3/1/05). Another group of donors subsequently pledged $717 million for immediate emergency relief alone (an unprecedented amount according to UN officials). Then there were public appeals that raised millions of dollars. Since then, more pledges have poured in, and by 2007 the total amount committed for the reconstruction of Aceh reached US$7.77 billion, of which donors have committed the largest share (42 percent), followed by non-government organizations-NGOs (30 percent), and the Government of Indonesia (29 percent).

Along with a temporary suspension of debt payments to Paris Club creditors, the financial assistance from donors has created effects on the Indonesian national budget and the balance-of-payment limited. The effect on Aceh itself, however, is very significant. Although the province accounts for only about 2 percent of the total Indonesian GDP, and the oil and gas sector that accounts for almost half of Aceh regional gross domestic product (RGDP) was not damaged by the disaster (Aceh's economy was reported shaved by 20 percent), the earthquake and the tsunami claimed heavy casualties, destroying infrastructure, settlements, schools, health centers, shops, and public buildings, left half a million people homeless, and dampened people's sense of optimism. The flow of international aid workers, along with supply and distribution constraints have sent food, housing and transportation prices soaring. Ironically, this has made the life of many local people more difficult as it creates an artificial economy in the region and a huge shock for the local economy, which will take a long time to recover.

Like in most disaster-recovery programs, the rebuilding process got off to a slow start, partly because of the huge scale of the catastrophe, and also because the first priority was getting food and basic shelter to survivors and helping them find jobs. This is understandable. A few years after the disaster, when the situation had calmed down and a special body known as Agency for the Rehabilitation and Reconstruction of Aceh-Nias (BRR), fully supported by the government, had begun its work in a full scale mode, progress was made. By December 2007, 83 percent of the total funds committed had been allocated to specific projects, and the Government (through BRR), has committed all its funds. Yet, such progress has been clouded by bureaucratic issues and other obstacles, raising concerns among donors and recipients. One of the major concerns pertains to the housing reconstruction activities. Not only have tens of thousands of people remained homeless, but those who were lucky enough to have homes had to suffer from a lack of supporting-infrastructure and poor access to basic services such as water, electricity, road, school, health centers, market, etc. All these were driven by the fact that the locations of the constructed houses were either not carefully planned, or the plan was not well implemented. While at the early stages many believed the disaster provided an excellent opportunity to re-build a new Aceh, the reality shows that such an opportunity has been clearly missed; those responsible were unable to make a virtue of necessity.

Compared to the newly built houses, many of the pre-tsunami houses were of better quality, environmentally more sustainable, more affordable, and more comfortable. Various reports reveal that some of the new houses are less safe and many of them have poorer access to the necessary infrastructure. Also, the official in-charge and some NGOs that participated in the reconstruction program did not really pay any attention to the socio-cultural and environmental context of the programs. The complexity and cultural sensitivity in housing and the links between the built environment and sustainable development are still not fully appreciated.[1] But lack of supporting infrastructure due to the unacceptable location of the built houses is of more immediate concern to many residents who are the victim of the disaster. Many areas continue to pose difficulties because of the inaccessibility to clean water, schools, health centers and even roads to market places. The major public health priorities of ensuring the availability of clean water, adequate sanitation, and

---

[1] A rare example was the reconstruction projects coordinated by the Government of Gujarat after the 2001 Earthquake. The project was conducted such that the process also strengthened local housing culture and building capacity by empowering local people through financial and technical assistance to manage themselves (Duyne 2006).

emergency food rations, should have not been too difficult to accomplish, had the housing reconstruction been conducted based on a more comprehensive plan that takes into account the location of such necessary supporting infrastructure. When some of this infrastructure is not in place, either due to the tsunami-related destruction or simply because it never existed there, careful coordination needs to be made to identify the infrastructure needs in each location and determine who will do what to meet them, and to determine the logistic for transporting and delivering the necessary goods and services. One of the most critical indicators to measure the accessibility to the supporting infrastructure is the distance between the built houses and the particular infrastructure. The problem becomes more complex when some infrastructure may be available (or planned to be built) in some areas but other necessary infrastructure is nowhere near those areas. Thus, a particular approach (model) needs to be used to find the optimal locational configuration of the houses to be built. This is the focus of this study.

Lack of a plan or the use of a bad plan led to an inefficient housing configuration. This made the whole reconstruction program far from cost-effective, the resulting outcome of which was a reduced welfare effect of the program.[2] The size of the welfare loss depends on the extent to which the realized location of the constructed houses and infrastructure deviates from the "optimal" configuration. A relevant question to ask is, therefore: what is the gap between the constructed houses and the "optimal" configuration? This is the main question addressed in this thesis. We attempt to approach the question by finding the "optimal" with respect to a particular set of areas or villages to include in the housing reconstruction, given the budget constraint and other physical restrictions (e.g., in the selection process we

[2] Another issue is the environmental impact. The poor design and construction quality of housing can have very significant environmental consequences such as the inadequacy of sanitation systems.

cannot select only a part of a village, it must be a full village, and the number of clusters should be less or equal to the number of selected villages). The key parameter used in the optimality measure is the ‘physical distance,’ implying that the transport costs and other obstacles to have access of the supporting infrastructure play a central role in determining what should be the site and the cluster configuration of the housing construction. To the extent that one of the main goals of the post-tsunami reconstruction program is to provide shelter and supporting infrastructure (drinking water, sanitation, drainage, electricity, schools and health centers) for the victims, the above question is highly relevant for policy making and evaluation. [3]

3.2 Socio-Economic Analysis of Aceh

3.2.1. Poverty

Poverty in Aceh increased slightly in the aftermath of the tsunami, from 28.4 percent in 2004 to 32.6 percent in 2005 (World Bank, January 2008.) The increase is relatively small given the extent of damage and destruction caused by the tsunami and reflects the beneficial effects of the initial reconstruction effort. This occurred against failing poverty levels in the rest of the country. In 2006, poverty fell to 26.5 percent which is below the pre-tsunami level, suggesting that the rise in the first year of tsunami-related poverty was temporary and reconstruction work most likely facilitated this decline. The poverty level in the rest of the country, however, went up in 2006. Nevertheless, poverty in Aceh remained significantly higher than in the rest of Indonesia.

[3] This work is mainly a computational exercise applied to the City of Banda Aceh. The objective function developed in this essay does not consider social and economic aspects that could be essential. We would need a greater interaction with the people living in the villages to define the utility of village and the cluster benefit.

Table 3.1. Poverty Level in Aceh 2004 - 2006

Poverty Level (%)	2004	2005	2006
Aceh Province	28.4	32.6	26.5
- Urban	17.6	20.4	14.7
- Rural	32.6	36.2	30.1
Indonesia	16.7	16.0	17.8

Source: BPS and World Bank calculation

Poverty in Aceh is predominantly a rural phenomenon, with over 30 percent of rural households living below the poverty line, compared to less than 15 percent of poor households in urban areas (The World Bank, January 2008). Common characteristics related to higher poverty levels are lower education levels, larger household size, female-headed households and households that work in agriculture. Despite the rapid socio-economy and political changes, the relationship between these characteristics and poverty remained relatively stable over the period, suggesting that underlying determinants of poverty were unchanged.

Aceh has experienced very low or negative growth rates for most of the past 30 years, lagging behind the rest of the country in most years. The key reason for this slower growth was the longstanding internal armed conflict affecting the province, although deficiencies in structural economics also contributed to the poor performance. The abundance of natural resources in Aceh, primarily large gas and oil reserves on Aceh’s east coast that resulting high GDP per capita, has not resulted in higher growth rates or lower poverty levels. In fact, the wealth of natural resource and the conflict resulted in a struggling economy, weaker government, low public services performance as well as some of the highest poverty levels in Indonesia.

3.2.2. Structure Economy

Aceh's economy depends heavily on the mining sector, which includes oil and gas extraction, quarrying activities of sand, gravel and stone, which accounted for 24.9 percent of GDP in 2006 (The World Bank, 2007).

Before the tsunami occurred, in 2003 this sector accounted for 36.1 percent of GDP, which is the largest in Aceh as shown in Table 3.2 and Figure 3.1. The manufacturing sector, which contributed 14.3 percent to the GDP in 2006, is directly related to the availability of cheap gas, however, due to the depletion of known gas reserves, many manufacturing units were closing. The decline in the mining and related manufacturing sectors was further accelerated by the impact of the tsunami. Conversely, the service sector has seen an expansion due to the post-tsunami reconstruction effort.

Table 3.2. Structure of Aceh Economy 2003-2006

Sector (%)	2003	2004	2005	2006
Agriculture and fisheries	17	20	21.4	21.2
Oil, Gas and Mining	36.1	30.4	26.2	24.9
Manufacturing Industries	20.2	18.3	15.9	14.3
Electricity and water supply	0.1	0.1	0.2	0.2
Building / construction	3.4	3.8	3.5	5.1
Trade, hotel and restaurants	11.2	12	14.3	15
Transport & communication	3.3	3.8	4.8	5.2
Banking & other financial	0.9	1.2	1.2	1.3
Services	7.8	10.4	12.7	12.9

Source: BPS (2002 = 100)

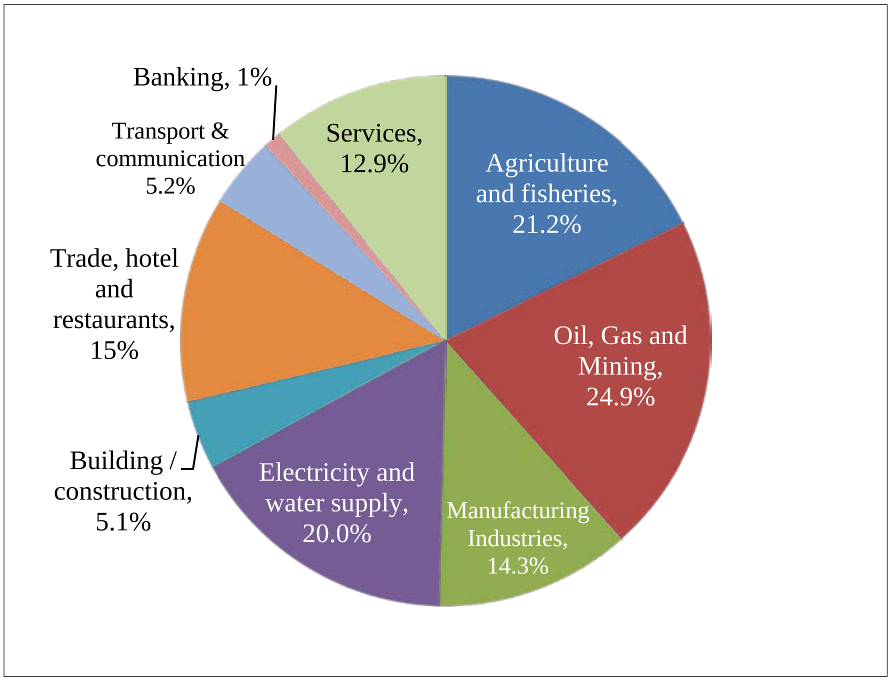


Figure 3.1. Aceh's Economy, 2006

3.2.3. Unemployment

The negative economic growth has contributed to continuing unemployment. The decline of Aceh's economy before and during the tsunami has contributed to the growing unemployment problem in the province. Limited growth in some sectors, such as agriculture or some services sectors, has not translated into significant employment generation. Unemployment increased from about 6 percent in 2000 to 12 percent in 2006 despite the large reconstruction effort after the tsunami that created significant job opportunities, as shown in Figure 3.2.

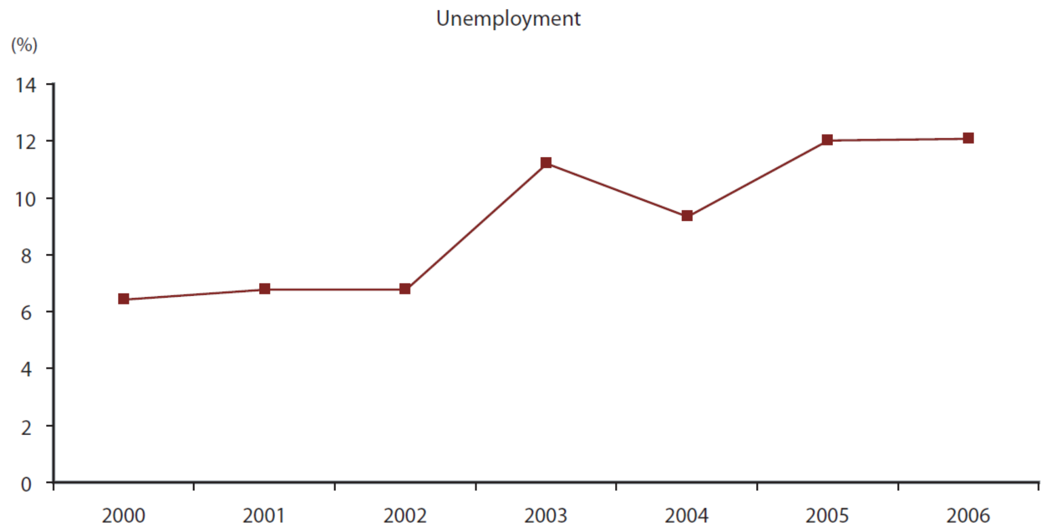


Figure 3.2. Unemployment in Aceh, 2000 – 2006

3.2.4. Economic Growth

Aceh's economic performance pre-tsunami was affected by both a decline in the gas reserves and a decade-long armed conflict between GAM (Free Aceh Movement) and the government. The decline in gas production has had a negative impact on industries that were dependent on the accessibility of inexpensive gas and situated close to the gas fields, such as chemicals, paper, or fertilizer. Growth in Aceh post-tsunami, especially in 2006 and 2007, has been dominated by sectors closely connected to the reconstruction effort, such as construction, trade and transport, while agriculture also showed positive growth.

Table 3.3. Economic Growth in Aceh, 2004 - 2008

Sector (%)	2004	2005	2006	2007	2008*
Agriculture, forestry & fisheries	6.0	-3.9	1.5	3.6	0.8
Mining and quarrying	-24.0	-22.6	-2.6	-21.6	-44.7
Oil and Gas	-24.4	-23.0	-4.3	-22.5	-47.0
Quarrying	7.3	0.8	78.8	2.0	-0.2
Manufacturing industries	-17.8	-22.3	-13.2	-10.1	-4.2
Oil and gas industry	-11.6	-26.2	-17.3	-16.7	-7.8
Non-oil and Gas Industry	-37.3	-5.1	1.1	8.6	3.6
Electricity, gas and water supply	19.5	-2.0	12.0	23.7	12.7
Construction	0.9	-16.1	48.4	13.9	-0.9
Trade, hotel and restaurants	-2.6	6.6	7.4	1.7	4.6
Transport & communication	3.6	14.4	10.9	10.9	1.4
Financial	19.4	-9.5	11.7	6.0	5.2
Services	20.1	9.7	4.4	14.3	1.2
GDP Aceh	-9.6	-10.1	1.6	-2.5	-8.3
GDP Aceh w/o oil & gas	1.8	1.2	7.7	7.0	1.9
GDP Indonesia	5.0	5.7	5.5	6.3	5.9
GDP Indonesia w/o oil & gas	6.0	6.6	6.1	6.9	6.4

Source: BPS and World Bank calculation. *Preliminary figures

As the reconstruction effort was coming to an end in 2009, economy growth has began to slow down and preliminary numbers for 2008 show a considerable deceleration in sectors formerly stimulated by the reconstruction effort. Aceh's economy, including oil and gas, declined by 8.3 percent in 2008, with the non-oil and gas economy growing slowly by 1.9 percent, well below 6.4 percent at the national level.

3.2.5. Housing and Infrastructure

Housing construction in 2006, two years after the Tsunami, has made significant progress with around 57,000 permanent houses completed in the end of the year, representing 50% of the overall housing reconstruction needs in Aceh and Nias

(BRR, December 2006). Additionally, 15,000 transitional houses were built. Although the numbers look promising, ironically, only 14 percent (65,000) Internally Displaced People (IDPs) out of 500,000 had been moved out of tents into transitional housings by the end of 2006.

Difficulties in the housing reconstruction affected many agencies, including the Earthquake and Tsunami Emergency Support Project (ETESP), whose housing program comprised about 6,000 newly reconstructed housing units (ADB, 2010.) This is considerably less than both the 14,000 units anticipated at ETESP project appraisal, and the 8,000 units envisioned in March 2006 when the housing program's overall target was downscaled. Essentially, unit cost increases due to price hikes and quality improvements, and implementation and land constraints account for this difference between the number of reconstructed units planned and that achieved.

Almost 750 permanent schools were built throughout 2006, supplemented by 379 temporary schools such that the vast majority of children in Aceh and Nias are in school (BRR, December 2006.)

A total of 324 health facilities in Aceh and Nias have been repaired or reconstructed in 2006 and efforts to improve the quality of available health services have begun through the training and capacity development of key health workers (BRR, December 2006).

In the infrastructure sector, 1,500 km of all types of roads in Aceh and Nias out of 3,000 km of impassable road had been built or repaired by the end of 2006. Furthermore, 158 bridges out of 1,620 arterial and minor bridges that were destroyed had been repaired (BRR, December 2006.)

3.3 Modeling Housing Location

Most post-disaster agency-driven housing reconstruction projects have a narrow technical approach (Barenstein and Pittet, 2007; Twigg 2002). Even organizations that in normal times are committed to sustainable development, in an emergency context often make technological choices without taking into account the sociocultural, environmental and economic implications (Barakat 2001; Duyne 2006). This is despite the fact that many efforts have been made to define principles and practices for the post-disaster housing sector (Sphere Project 2004).

The experience of the housing reconstruction program in Aceh is not much different. First of all, one has to realize the extent of damage. The effects of the Tsunami were evident up to 4 km inland from the Aceh coast. Many of the buildings within 2 km of the coastline were washed away and demolished. A large amount of water swept and washed through the houses and roads. Figure 3.3 shows the map of Banda Aceh before and after the tsunami.

Figure 3.3. Banda Aceh Before and After Tsunami

Source: DigitalGlobe's QuickBird Natural Color Image, showing Banda Aceh shoreline before Tsunami, June 23, 2004 (left) and the missing shoreline after Tsunami, December 28, 2004.

The fatalities in the densely populated Banda Aceh, the capital city of Aceh province, were estimated to be over 30,000, and almost 18,000 people were internally displaced during the early post-disaster period. But the impact of the tsunami was greatest in the two western coastal kecamatan (districts), Meuraksa and Kuta Raja. From these two areas alone there were more than ten thousand deaths or missing persons. Another area severely damaged was the south-western kecamatan of Jaya Baru, where 5,966 deaths and missing persons were reported. This is despite the fact that this district is far from the coast. The next most severely damaged (partially damaged) areas are: Baiturrahman, Syiah Kuala, and Kuta Alam, all of which are located in the district of Kota Banda Aceh. Therefore, the city of Banda Aceh is to be the area of analysis (see Figure 3.4). The three districts that escaped from the tsunami are: Ulee Kareng, Lueng Bata, and Banda Raya. Looking at the village level, about 41 villages were totally damaged (52.8 %), 8 partially damaged (3.3 %), and 40 villages remained in good condition (43.9%).

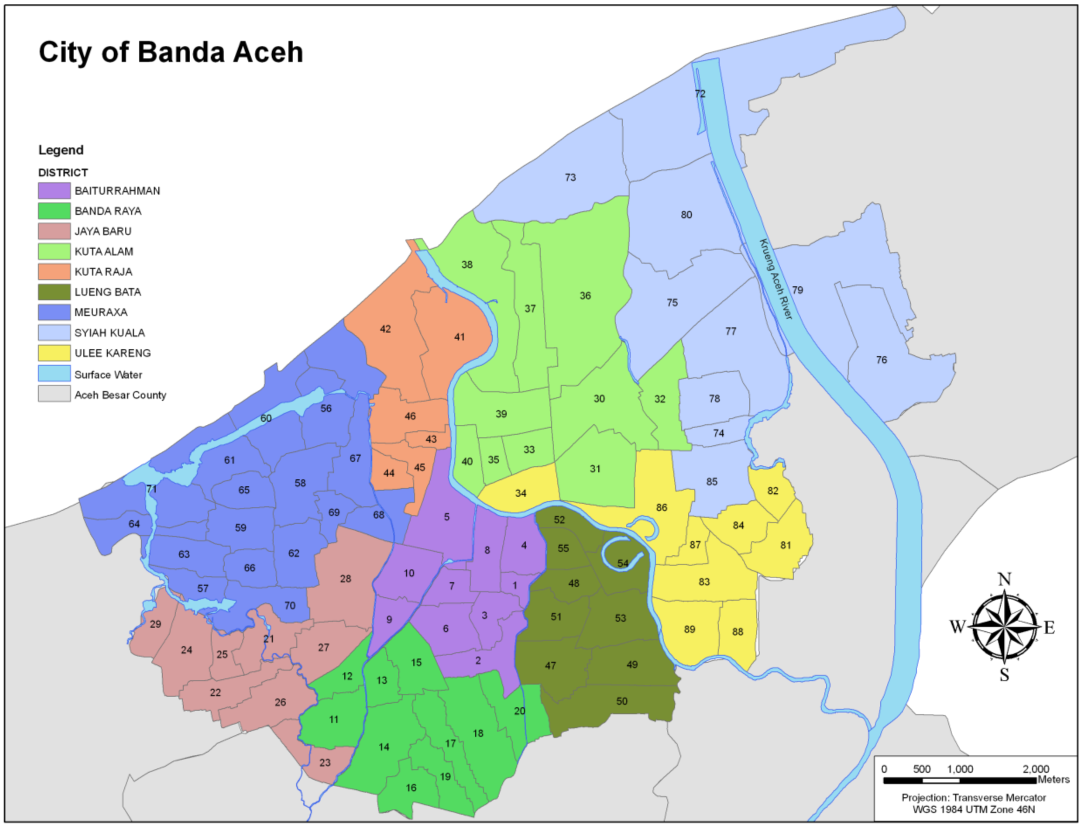


Figure 3.4. Study Area: City of Banda Aceh

In this study, 89 villages in Banda Aceh city are used as the unit of observation. We represent the City of Banda Aceh as a planar graph in a two dimension space, $G = (V, E)$. The set of village V corresponds to the villages, i.e. each village is a member of s_v of V:

$$V = \{s_v \mid v = 1, 2, \ldots, n\}; \quad (1)$$

where n is the number of villages. Each village is evaluated based on its utility u_v and a cost c_v. Since our goal is to find the optimal locational configuration of villages where housing construction is to take place, there is a possibility that more than one village adjacent to others is selected; then a cluster is therefore formed. In the next sub-chapters we introduce variations of knapsack problems and compare the results.

3.3.1 The Basic Knapsack Model

The Basic Knapsack Model is formulated as follows:

$$\text{Maximize } \sum_{v} u_v x_v \tag{2}$$

$$\text{s.t. } \sum_{v} c_v x_v \leq M \tag{3}$$

$$x_v \in \{0,1\} \qquad \forall v \in V \tag{4}$$

where u_v is the utility for each village v, x_v is the binary variable associated with each village v (1 reflects the case where the village is selected, 0 is otherwise), c_v is the cost of housing reconstruction in village v, and M is the housing reconstruction budget.

To the extent that the optimal location of houses to be built is determined by not only the extent of the damage but also the availability and location of supporting infrastructure, the variables included in the utility to be maximized ought to include each type of socio-economic infrastructure (e.g., health, education, clean water). It is assumed that the system of villages and clusters is additive for computational convenience. The village utility function used in our model is:

$$u_v = w_1\left(\frac{area}{DWater}\right) + w_2.Health + w_3.DEdu + w_4.WaterNet \tag{5}$$

where u_v = Utility of village; w_j = proportion of utility where w_j = 0.25, j=1,2,3,4; *area* = area of a village; *DWater* = distance to water body from the center of village v; *Health* = number of health centers (hospitals) in a village v; there are a total of 29 health centers (hospitals) in Kota Banda Aceh; *Edu* = distance of a school to the village; there are a total of 128 elementary schools in Kota Banda Aceh; *WaterNet* = the existence of drinking pipe network diameter 200-600 mm (1 if pipe line exist in village *v*, otherwise 0). To measure the distance of center of village *v* to the nearest water body, we use ESRI ArcGIS *Proximity Analysis*.

Azzaino et al. (2002) developed a framework using individual land parcels simulated via hydrologic models including the ratio of parcel size divided by the parcels distance to the water intakes. We use a similar approach for calculating the ratio area of village *v* divided by the distance of village *v* to the water body, *DWater*. The distance to the water body plays an important role in the tsunami affected area, as it is one of the main water sources for the villagers during the recovery period. Although most water sources were contaminated by the tsunami, Vithanage (2009) and Srinivas (2007) stated that the majority of streams and rivers in the post-tsunami affected area are expected to be naturally flushed clean over a shorter period of time, compared to the longer cleansing time and the more difficult to remedy contaminated ground water due to saltwater intrusion, sewage, debris and hazardous materials. In addition, the overall water quality of the main river, Krueng Aceh, is relatively good quality, only small sections can be considered moderately to fairly polluted with organic material (ESP, 2007).

Byleveld et al. (2005) explain that during the emergency period, the first several months after the tsunami, NGO personnel built *water points*, which drew water from a main pipe across the Krueng Aceh River where the water was treated by coagulation, filtration, and disinfection; then distributed to the villagers. This act was much appreciated by the villagers as they need clean water for their daily use. Thus we include the *WaterNet* attribute, to indicate the existence of a main water pipe in Equation (5).

In the village utility calculation, attributes *DWater*, *area* and *DEdu* were normalized using the following formula:

$$N\alpha_v = \frac{(\alpha_v - \alpha^{\min})}{(\alpha^{\max} - \alpha^{\min})} \tag{6}$$

Based on Equation (6), a village with the lowest attribute value has a normalized score of zero, and a village with the highest attribute score has a normalized score of one. Village #21 with normalized distance to water of zero, would make (*area*/*DWater*) undefined. Therefore, a small but positive normalized distance $N\alpha_{21}$=0.0035 was assigned to village #21.

In selecting the optimal location of villages and clusters, the officials in charge have to work within the given budget. Although in general the budget size is not the most binding constraint, largely due to the generous commitments by many institutions, foreign and domestic alike, the allocated fund for housing reconstruction is fairly fixed. More significantly, the magnitude of the damage was so large that any estimates of the costs of reconstruction needed to be adjusted upward. The numbers were simply staggering. Estimates by FAO and ADB suggest that the housing construction requires 1 million tons of cement, 3.6 million m^3 of sand, 1.1 billion fired clay bricks, 508,000 m^3 of concrete blocks, 87,000 m^3 of plywood, 370,000 m^3 of sawn timber and 945,000 m^3 of fuel wood for brick kiln firing (ADB, 2006). The environmental costs could also be high. According to UNEP (2007), to meet the need for fuel wood for brick making alone, "about 10,000 hectares of forest would have to be logged. These estimates are limited to the immediate plans to build 120,000 houses and do not include the materials required to construct other types of buildings, such as schools, mosques, hospitals and commercial buildings, or to modify houses after they have been completed."

BRR and Partners (2006) determined that average cost to rebuild a 36 square-meter house is approximately US$4,000. A total of 38,228 houses are to be constructed or rebuilt in the City of Banda Aceh.

It is important to point out, however, that in some cases costly reconstruction is unnecessary and undesirable. There are cases where the pre-disaster built houses are less expensive but also more favorable from the socio-economic and cultural perspectives, because human settlements reflect peoples' history and cultural identity and the built-in environment is usually shaped over long periods of time before finding its 'optimal' equilibrium. Demolishing reparable houses can be also unacceptable from the environmental point of view considering that building materials (wood) and technologies (burning, energy-waste) used to construct the new houses may detract valuable resources to meet other requirements.[4]

The initial budget 30 percent of total cost is given. We use CPLEX (ILOG, 2007) to construct the model and the result is visualized in Figure 3.5 as follows:

[4] Massive cutting of trees needed to clear the land is another environmental cost often undermined by builders and contractors.

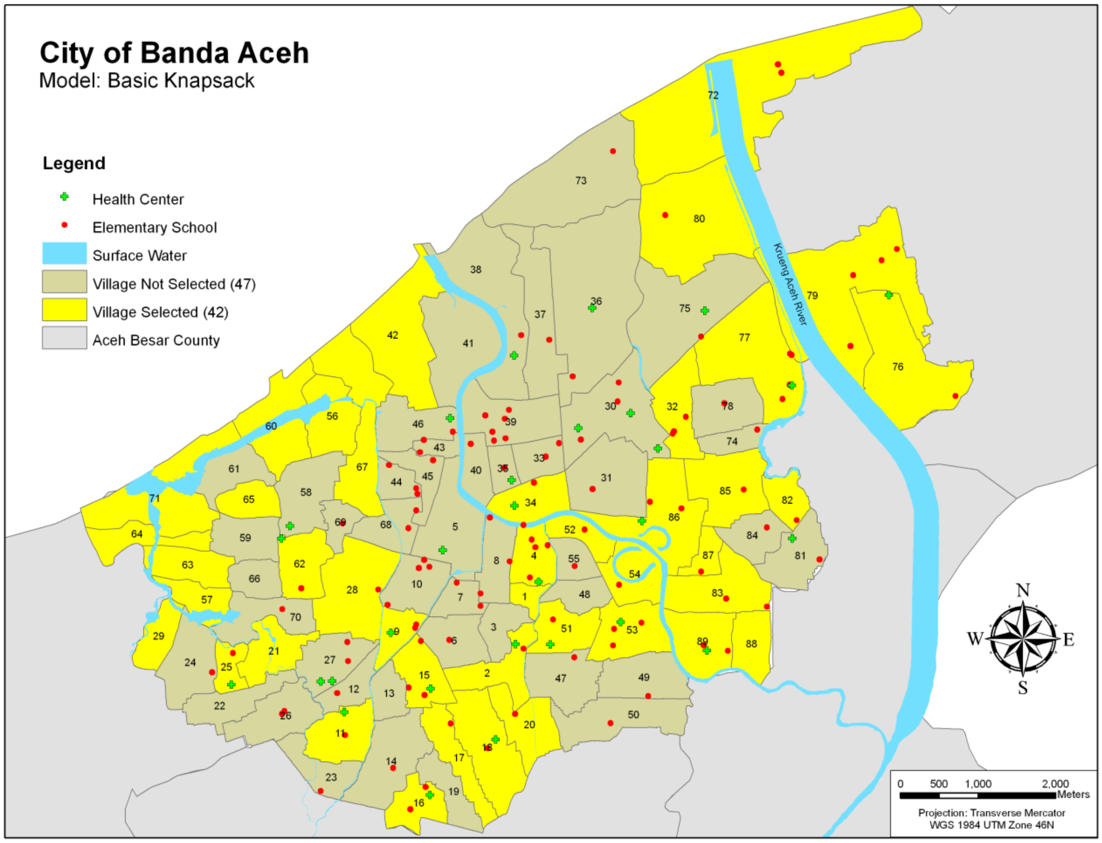


Figure 3.5. Locational Configuration of Housing Reconstruction: Basic Knapsack

Figure 3.5 shows the optimal solution for the Basic Knapsack Model. The yellow areas indicate the selected villages (42) that would be included in the first stage of housing construction. From the budget of Rp. 387,631,920,000, the total cost to rebuild housing in these 42 villages is Rp. 387,415,600,000. The construction will benefit 110,558 residents in the selected villages. Notice that in general, with the exception of the area in the north-east of Banda Aceh, most selected villages are relatively smaller villages than those in the non-selected areas. This may look questionable because in the objective function to be maximized includes the ratio between the size of the area and the distance from river. It turns out that the explanation rests on the distance, i.e., the denominator. Even if there are several large villages on the eastern part of the city and there is also a relatively large river in that area (Banjir Kanal Krueng Aceh), they are not selected because the distance of some of the large villages to the river (measured from the middle point of the village area) is too far such that the value of the ratio, thus the resulting utility, is relatively small. On the other hand, in many of the smaller size villages in the western part of Banda Aceh the rivers (Krueng Aceh and several water bodies, i.e., lakes, swamps) pass through the villages such that the distance from each of the villages to these rivers is relatively small, so that the resulting utility is large and more likely maximized.

In reality, Steinberg (2007) estimated that 50 percent of housing needed in the City of Banda Aceh was built in all the villages in the City and was accomplished after 2.5 years of the tsunami, with a great inconsistency in the quality of housing structure. In addition, the village of Punge Jurong (population 2,300 in 2007), prior to tsunami 6,000) needed approximately 2,000 houses, however by April 2007, two and half years after the disaster, only 550 houses have been built by the Indonesian Red Cross, 248 by UN Habitat, and 226 by BRR (Loomis et all., 2007). There are also concerns about the great disparity in the quality of construction of some of the replacement homes that have been built. Some of those made homeless by the tsunami have received what amount to plywood shacks for homes, while others will receive solid brick houses with plastered walls and tiled floors. Furthermore, houses were constructed in all villages including areas with damaged residential infrastructure and community facilities, and with clean water source no longer available.

In the Basic Knapsack model, it is assumed there is no clustering force, although it does not necessarily mean that it will not result in clustering. The resulting cluster, if any, is due to the physical location (proximity) of the selected villages, not because of the imposed weight. The following sections deal with a set of alternative models that will include clustering effects. The results of these models are to be compared with the Basic Knapsack results.

3.3.2 The Basic Clustered Knapsack Model

As discussed earlier, the actual locational configuration of constructed houses in Banda Aceh is far from optimal, as many of them are scattered throughout the city without considering whether there exists adequate residential infrastructure, community facilities, or other basic needs such as access to clean water, schools and health services. It is therefore of interest to determine the maximum number of clusters in the optimization system. The Basic Clustered Knapsack Model is developed with this issue in mind.

Having the largest number of houses to be built is not always the most preferred scenario, as it may not be the one that will provide shelter to the largest number of people affected by the tsunami. At the end of the day, the officials in charge should be more concerned with the latter. It is in this context, in evaluating the results of each alternative scenario the study will not only look at the number of clusters and villages where houses are to be built but also put the focus on the number of people that can be sheltered by the program.

Let C be the set of m clusters in Banda Aceh, where each cluster k is a subset of V:

$$C = \{ C_k : C_k \subseteq V \mid k = 1, 2, \ldots, m\}; \tag{7}$$

Each cluster C_k is characterized by a utility u_k. The objective is to find a set of nodes (villages) $W \subseteq V$ such that the sum of cost c_v is less or equal to the given budget M. The Basic Clustered Knapsack model can be formulated as follows:

$$\text{Maximize } \sum_k u_k y_k \tag{8}$$

$$\text{s.t. } \sum_v c_v x_v \leq M \tag{9}$$

$$y_k \leq x_v \qquad \forall k, v \in C_k \tag{10}$$

$$x_v \in \{0,1\} \qquad \forall v \in V \tag{11}$$

$$y_k \geq 0 \qquad \forall C_k \tag{12}$$

where u_k is the utility for each cluster k, y_k is the binary variable associated with cluster C_k (1 if all villages in the cluster is selected, 0 otherwise), x_v is the binary variable associated with each village v, c_v is the cost of housing reconstruction in village v, and M is the housing reconstruction budget.

The cluster utility u_k is defined as follows:

$$u_k = \sum_{v \in C_k} u_v \tag{13}$$

where u_v is similarly defined as in Equation (5).

3.3.3 Model Variations Based on Cluster Type

We introduce two types of cluster types: Neighbor and Edge. In *Neighbor*, each village induces a cluster; the cluster is made out of the village itself and all the villages adjacent to it. In *Edge*, a cluster corresponds to a pair of adjacent villages. Therefore all the clusters have size of two. The *Edge* type may be a more natural way of modeling the notion of a cluster.

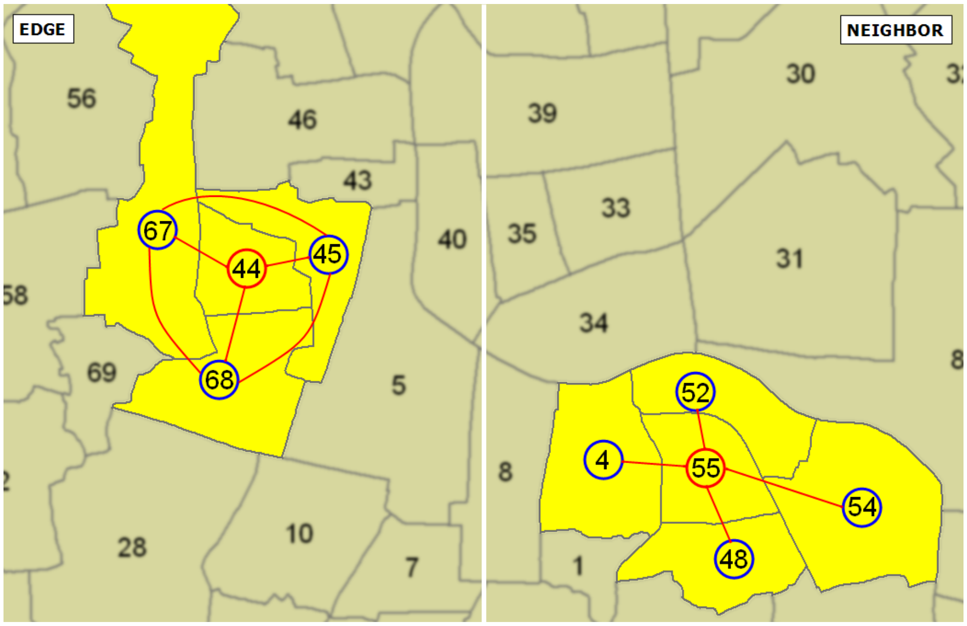


Figure 3.6. Cluster Type: Edge and Neighbor

Figure 3.6 shows the spatial differences between Edge and Neighbor cluster types. The right part of the illustration shows Neighbor types where cluster #55 is made out of village #55, 4, 48, 52 and 54. The left part of the illustration shows several Edge types: village #44-45, 44-67, 44-68, 45-67, and 45-68.

The optimal solution of Basic Clustered Knapsack Model is visualized in the following graph and table.

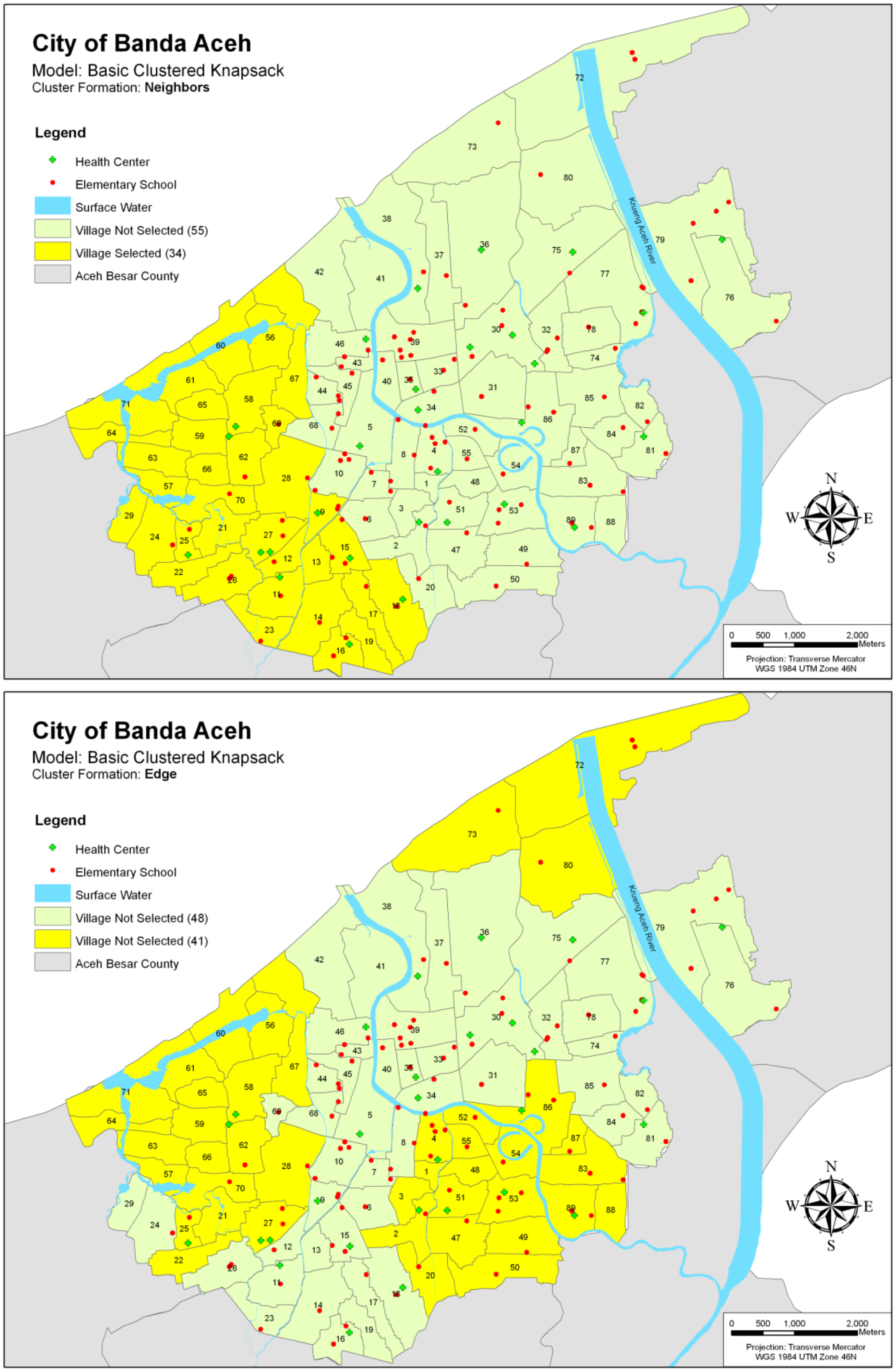


Figure 3.7. Locational Configuration of Housing Reconstruction: Basic Clustered Knapsack, Cluster Type: Neighbor and Edge

Table 3.4. The Basic Knapsack versus Basic Clustered Knapsack

	Basic Knapsack	Basic Clustered Knapsack	
		Neighbor	Edge
Total Cost (mill Rp.)	1,292,106	1,292,106	1,292,106
Budget ratio, a	0.3	0.3	0.3
Budget, M (mill Rp)	387,632	387,632	387,632
Parameter ε	-	1	1
Total Utility of Selected Village	70.82	55.18	64.20
Total Utility	70.82	267.64	288.37
Total Cost (mill Rp)	387,416	382,447	387,078
x_v, village selected	42	34	41
y_k, cluster selected	3	28	22
Edges selected	NA	NA	89
Pop2007 included	110,558	45,602	72,498

Table 3.4 shows the optimal solution results for the Basic Clustered Knapsack Model. When the cluster type is Neighbor, the number of villages selected is only 34 villages which consist of 45,602 residents. This is significantly smaller than the Edge cluster type that includes 41 villages with 72,498 people. Looking at Figure 3.7, it turns out those 34 selected villages in the Neighbor-cluster type form one cluster in the west part of the city, while the Edge type forms three clusters throughout the city. One might explain that when the cluster utility u_k is defined as the unit of measurement, the cost to acquire a single cluster in the Neighbor type is more expensive than the Edge type, thus it forces the model to select its nearest cluster until the budget is exhausted.

3.3.4 The Modified Clustered Knapsack Model

The objective function of the Clustered Knapsack Model is restated below:

$$\text{Maximize } \sum_{v} u_v x_v + \sum_{k} u_k y_k \qquad (14)$$

Note that there is a possibility of double counting utility of the objective function. To avoid double counting of the utility, we replace the second term of Equation (14) with the following equation:

$$\sum_{k} \varepsilon \prod_{v \in C_k} x_v \tag{15}$$

Parameter ε is a weight to quantify the benefit of utility of producing a cluster. When all of the items in cluster k are selected, the product of x_v equals to 1, on the other hand if one or more items in cluster k are not selected, then the product of x_v equals to 0. Normally, using a product of variables in the MIP would cause the program to become nonlinear. However, we can still use the variables y_k introduced previously to indicate whether all nodes in cluster k have been selected when under the assumption that we have found an optimal solution that would push each value of y_k to the correct value. Thus Equation (15) is replaced with the following equation:

$$\sum_{k} \varepsilon y_k \tag{16}$$

The objective function in Equation (14) therefore becomes:

$$\text{Maximize } \sum_{v} u_v x_v + \sum_{k} \varepsilon y_k \tag{17}$$

$$\text{s.t.} \quad \sum_{v} c_v x_v \le M \tag{18}$$

$$y_k \le x_v;\ \forall k,\ v \in C_k \tag{19}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{20}$$

$$y_k \ge 0;\ \forall C_k \tag{21}$$

We refer to Equation (17) as the **Modified Clustered Knapsack Model** (MCKM). When ε is equal to zero, the objective function in Equation (17) is equal to the Basic Knapsack in Equation (2). Thus we set the initial value of ε = 0.01. With all other parameters fixed, we vary the value of ε by increasing it in steps 0.01 (i.e. 0.01, 0.02, 0.03…) until $\varepsilon = 1000$. We then identify the value of ε that causes a change in the model results, i.e. that causes the solution to go from mainly unclustered to gradually more clustered and at some given ε the solution will be highly clustered. We run the MCKM problem in two different cluster types: Neighbor and Edge. The optimal solution is presented in Table 3.5, Table 3.6, Figure 3.8 and Figure 3.9 as follow:

Table 3.5. Results of Modified Clustered Knapsack Model (MCKM) - Neighbor

Budget, M (mill Rp)	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632
Parameter, ε	0.03	0.12	0.2	0.26	0.4	0.6	0.61	0.62	3.1
Total Utility of Selected Village	70.70	70.24	69.86	68.58	67.01	65.22	64.02	63.40	60.30
Total Utility	70.94	71.68	72.66	73.52	76.21	80.82	81.10	81.38	153.30
Total Cost (mill Rp)	387,618	387,348	386,571	387,416	387,382	386,199	386,773	387,585	386,773
x_v, village selected	41	43	42	42	44	46	45	45	42
y_k, cluster selected	8	12	14	19	23	26	28	29	30
Edges selected	NA	NA	NA	NA	NA	NA	NA	NA	NA
Pop2007 included	116,141	110,580	115,532	119,507	113,060	93,204	85,283	83,514	71,878

Table 3.6. Results of Modified Clustered Knapsack Model (MCKM) - Edge

Budget, M (mill Rp)	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632	387,632
Parameter, ε	0.02	0.04	0.05	0.06	0.07	0.08	0.13	0.29	0.33	0.45	1.38	92.21
Total Utility of Selected Village	70.75	70.57	70.52	70.17	70.04	69.41	68.79	67.38	66.41	65.52	61.39	61.29
Total Utility	71.91	73.09	73.72	74.37	75.08	75.81	79.84	93.48	97.10	108.27	196.63	9097.87
Total Cost (mill Rp)	386,976	387,280	387,551	387,618	387,416	387,179	387,078	387,551	387,213	386,165	387,145	387,517
x_v, village selected	42	43	43	43	44	46	46	48	48	49	44	44
y_k, cluster selected	3	4	4	5	9	8	17	16	21	21	28	29
Edges selected	58	63	64	70	72	80	85	90	93	95	98	98
Pop2007 included	114,567	111,374	108,792	109,247	108,037	101,557	97,889	94,190	89,626	93,938	83,386	86,889

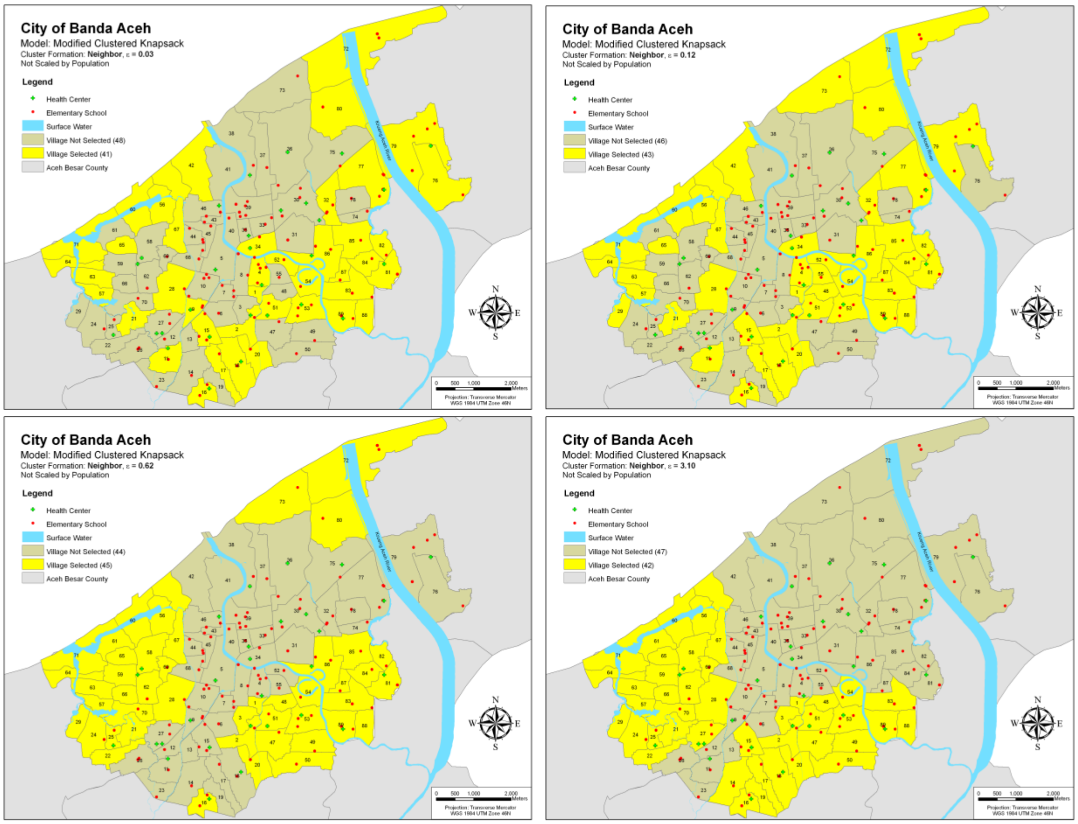


Figure 3.8. Locational Configuration of Housing Reconstruction: MCKM - Neighbor

The solution changes when parameter ε is equal to 0.03, 0.12, 0.20, 0.26, 0.40, 0.60, 0.61, 0.62, and 3.10. The lowest value of ε that changes the solution is 0.03, which means that when ε = 0.01, or ε = 0.02, the solution is the same as ε = 0. Similarly, when ε = 0.04, 0.05… 0.11, the solution is the same as ε = 0.12.

Figure 3.8 shows optimal solutions for the first two lower and upper extreme of multipier ε values, i.e. 0.03, 0.12, 0.62 and 3.10.

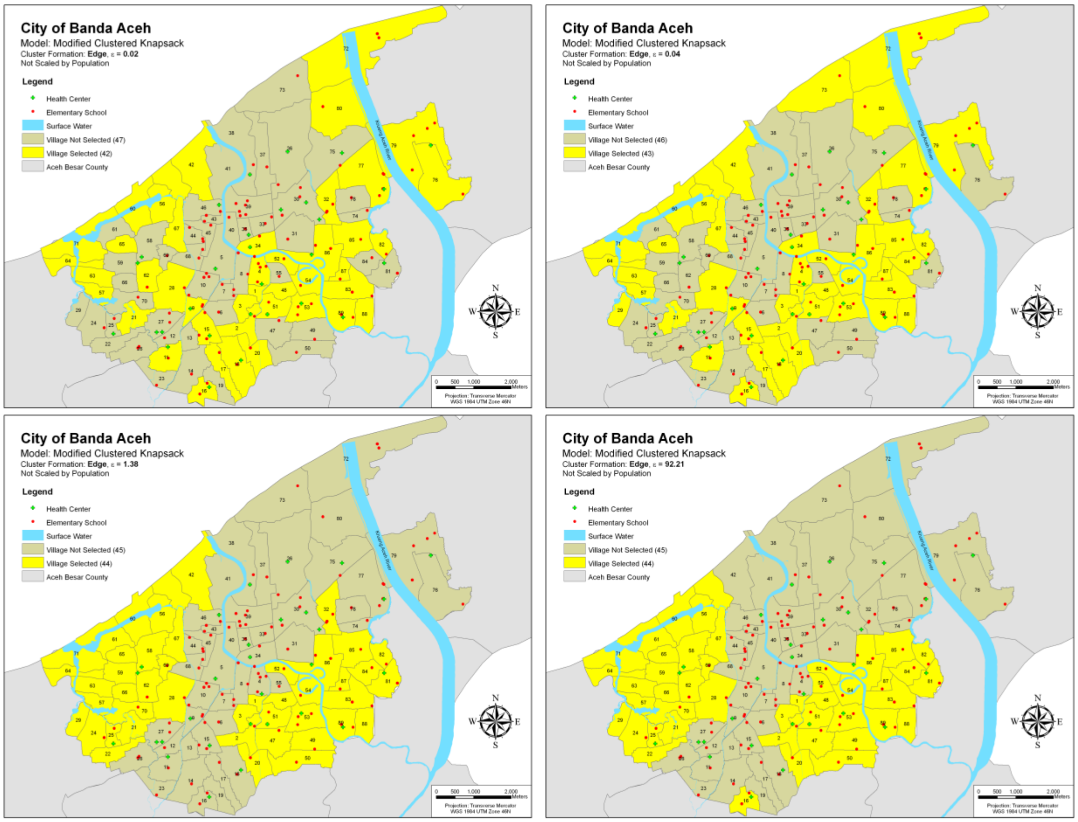


Figure 3.9. Locational Configuration of Housing Reconstruction: MCKM - Edge

The solution changes when parameter ε is equal to 0.02, 0.04, 0.05, 0.06, 0.07, 0.08, 0.13, 0.29, 0.33, 0.45, 1.38, and 92.21. The lowest value of ε that changes the solution is 0.02, which means that when ε = 0.01 the solution is the same as ε = 0. Similarly, when ε = 0.14, 0.15… 0.28, the solution is the same as ε = 0.13.

Figure 3.9 shows optimal solutions for the first two lower and two upper extreme of parameter ε values, i.e. 0.02, 0.04, 1.38 and 92.21. Both MCKM Neighbor and Edge problems results show that when ε is small the solution will mainly be unclustered, the higher ε the solution will gradually more clustered.

The number of people covered in the first MCKM Neighbor scenarios where ε = 0.03, 0.12, 0.2, 0.26 ad 0.4 (which basically means ε = 0.01 – 0.59) are between

110,580 – 119,507 people, which are higher than the Basic Knapsack model (110,558 people). The number of village selected are either the same (42 village at $\varepsilon = 0.20$) or higher (43 – 44 villages). Note that when $\varepsilon = 20$, the villages selected are not necessarily the same with the villages selected in the Basic Knapsack Model. Similarly, in the MCKM Edge scenarios where $\varepsilon = 0.01 – 0.04$, 43 villages are selected covering between 113,374 to 114,567 people compared to 42 village with 110,558 in the Basic Knapsack Model.

As mentioned before, decision makers will not only look at how many villages or clusters can be covered, but should be more concerned with how many villagers will benefit from the reconstruction program.

3.4 Conclusion

Ultimately, the tsunami reconstruction efforts needed to satisfy three separate entities, each with different, and sometimes conflicting, sets of expectations. They include the victims of the tsunami, the Government, and donor agencies. In providing housing assistance to the victims of the disaster, rules and standards that are acceptable to both the Government of Indonesia and donors need to be developed. However, the government guidelines were formulated over a prolonged period with multiple modifications. Given the magnitude of the disaster and unpredicted coordination issues that arose during the first 18 months of the effort, it was not until March 2006, nearly 15 months after the tsunami, that draft guidelines for identifying beneficiaries and determining the amount and type of assistance were released.[5] By

[5] Initially, the Government assigned the tasks of strategy formulation and administration of the reconstruction effort jointly to the National Development Planning Agency (BAPPENAS) and the Ministry of Public Works (MPW). However, following establishment of the Rehabilitation and Reconstruction Agency (BRR) in April 2005, strategy formulation and administration became BRR's responsibilities.

that time, many donor agencies and NGOs had already begun their first projects and housing constructions.

As many NGOs had established their presence in the areas destroyed by the disaster by providing emergency medical supplies, potable water, food, and emergency shelter, victims of the tsunami pushed them to speed up the process of building permanent shelter. In some cases, the local communities who are affected by the disaster would hold meetings with multiple donors, and accept assistance from whichever agency was able to provide permanent housing at the earliest possible time. This issue complicates the overall planning of the reconstruction initiatives in Aceh.

Furthermore, the socio-economic effect on the Aceh economy is very significant. The incoming flow of massive numbers of humanitarian aid workers, along with supplies and distribution of goods caused food, housing and transportation prices increase dramatically. Unfortunately, it creates an artificial economy in the region and a huge shock for the local economy, which will take a long time to recover. Also, the socio-cultural and environmental context of the rebuilding effort was not considered seriously. The complexity and cultural sensitivity in housing and the links between the built environment and sustainable development are still not fully appreciated. Many areas continue to pose difficulties due to inaccessibility to clean water, schools, health centers and even roads to market places. The problem becomes more complex when emergency or permanent housing is built while necessary residential infrastructure is nowhere near the area.

The main question addressed in this thesis is to find the gap between the accomplished housing construction and the optimal configuration. We attempt to approach the question by finding the “optimal” set of areas or villages to include in the housing reconstruction, given the budget constraint and other physical restrictions (e.g., in the selection process we cannot select only a part of a village, it must be a full village, and that the number of cluster should be less than the number of selected villages).

Two and half years after the 2004 tsunami, it is estimated that only 50 percent of housing constructions needed is completed throughout Banda Aceh. Superimposing this accomplishment with our baseline scenario, we clearly see the difference between them. Housing was constructed in all villages including areas with residential infrastructure, community facilities and water source not readily available, while our model results considered available residential facilities and infrastructures that are available although they might need renovations.

In this chapter we have considered three knapsack problems: the Basic Knapsack, the Basic Clustered Knapsack, and the Modified Clustered Knapsack models. In the Basic Knapsack Model, with a budget 30 percent of total cost, we show that the result selects 42 villages out of 89 villages, which will benefit 110,558 people or about 47 percent of total 233,218 residents in the City of Banda Aceh. Although there is neither force of clustering nor weight of cluster benefit, the result in the Basic Knapsack model show that three clusters were formed. This is simply due to the proximity or adjacency of selected villages.

In the Basic Clustered Knapsack model where the smallest unit of measurement in the objective function is the utility of a clustered village u_k (instead of utility of village u_v), the optimal solutions for both Neighbor and Edge-cluster types show that the selected villages were formed as one and two big clusters, respectively. The total number of villages selected and the total population benefiting from the solution, however, are less than the solution from the Basic Knapsack model.

In the last model, the Modified Clustered Knapsack (MCKM), Neighbor and Edge-cluster types, the optimal solutions show that at the lower parameter ε the selected villages are relatively not clustered. Meanwhile, when ε increases the solution will gradually form bigger cluster(s). Furthermore, the numbers of villages selected and total population benefit from the solution are higher than those in the Basic Knapsack Model.

On the other hand, the results of the study also reveal that having the largest number of villages covered in the reconstruction program is not always the most preferred scenario, as it may not be the one that will provide shelters to the largest number of population affected by the tsunami, as shown in MCKM, Edge-cluster type where parameter ε equals to 0.45 there are 49 villages selected facilitating only 93,938 people.

All problems had the same reconstruction budget of Rp. 387,631,920,000. Eight villages, Bitai (#21), Punge Blangcut (28), Alue Deah Tengoh (56), Asoenanggro (57), Deah Baro (60), Kampong Pie (64), Lambung (65), Ulee Lheue (71), with a reconstruction cost of Rp. 64,051,000,000 covering 5,322 residents, were common to all optimal solutions as shown in Figure 3.10. We consider these villages as high priority villages to be included in the first stage of the housing reconstruction.

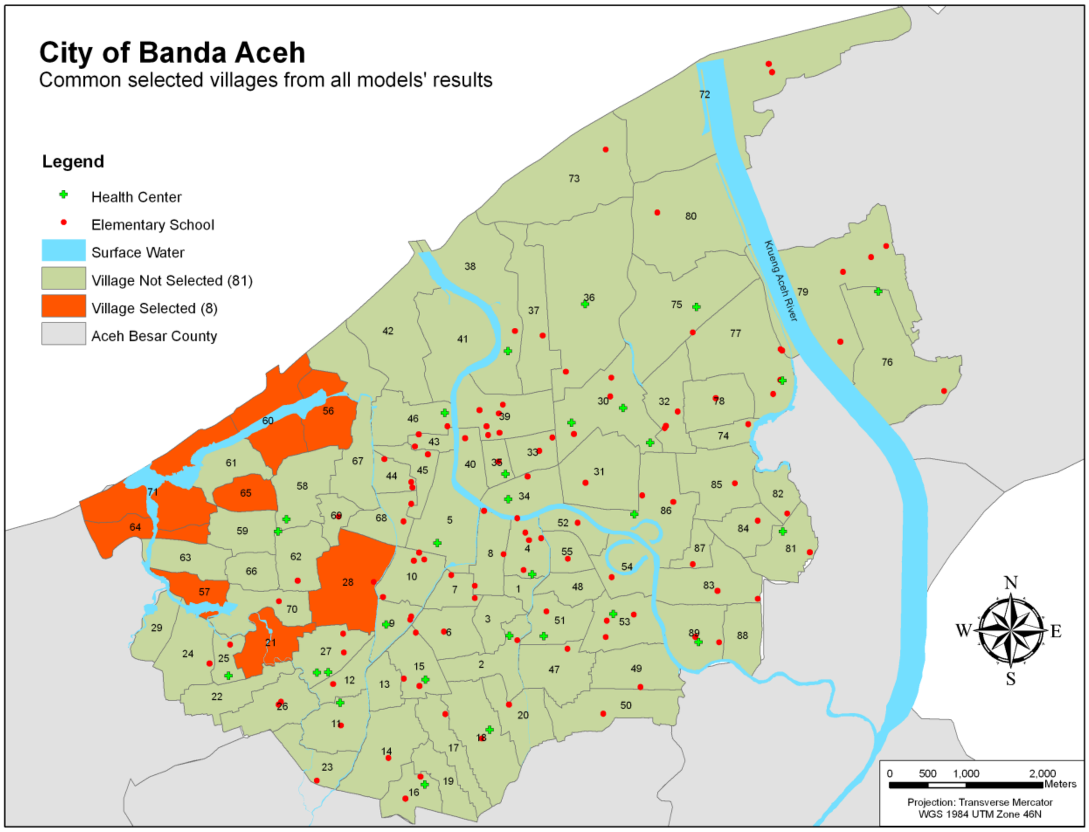


Figure 3.10. High Priority Villages to be Included in the Housing Reconstruction

Finally, we note that a number of parameters in the objective function were considered as a computational exercise. In order to generate a practical approach, a more comprehensive social and economic study is required. After the economic boom from reconstruction effort in Aceh due to the 2004 tsunami and earthquake, growth in the non-oil and gas sectors in Aceh has dropped significantly as the rebuilding activities draw to an end. The economic growth in the Aceh province is shown in Table 3.7.

Table 3.7. Economic Growth in Aceh, 2004 - 2009

Sector, growth (%)	2004	2005	2006	2007	2008	2009*
Agriculture, forestry & fisheries	6.00	-3.90	1.52	3.62	0.81	2.56
Mining and quarrying	-24.00	-22.60	-2.58	-21.10	-27.31	-47.28
Manufacturing industries	-17.80	-22.30	-13.18	-10.10	-7.73	-7.85
Electricity, gas & water supply	19.50	-2.00	12.06	23.70	12.73	6.45
Construction	0.90	-16.10	48.41	13.93	-0.85	13.79
Trade, hotel and restaurants	-2.60	6.60	7.41	1.70	4.59	4.94
Transport & communication	3.60	14.40	10.99	10.95	1.38	4.88
Financial	19.40	-9.50	11.77	6.02	5.16	7.83
Services	20.10	9.70	4.41	14.30	1.21	4.02
GDP	-9.60	-10.10	1.56	-2.36	-5.27	-5.51
GDP w/o oil & gas	1.80	1.20	7.70	7.23	1.88	3.97
GDP Indonesia	5.03	5.69	5.50	6.28	6.06	5.69
GDP Indonesia w/o oil & gas	5.97	6.57	6.11	6.87	6.52	6.20

*= Preliminary figures. Source: Statistics Indonesia (BPS)

As many of the reconstruction agencies, the NGOs and the Government of Indonesia considerably scaled down their operations in the province, growth in Aceh's non-oil and gas economy declined sharply as the reconstruction effort slows down. GDP growth fell to 1.88 percent in 2008 and 3.97 percent in 2009, far below the national growth rate of 6.52 percent in 2008 and 6.20 percent in 2009 for the non-oil and gas economy. Sectors related to the reconstruction effort that had led growth in Aceh since 2005 registered low or negative growth rates. The construction sector dropped to -0.85 percent in 2008, far below the 13.93 percent growth rate in 2007. Meanwhile, the oil and gas sector continues to decline, as a result of the rapid depletion of gas reserves, with production declining since 2008.

The agricultural sector, which was expected to become an engine of growth after the reconstruction phase, has been disappointing. After increasing by 3.62 percent in 2007, similar to the national growth rate for agriculture, it dropped to 0.81 percent growth in 2008, and although raised to 2.56 percent growth in 2009, it is less than the 3.62 percent in 2007. To be fair, however, not all of the disappointing performance is due to policy. The slow growth in the agricultural sector also stems from adverse rainfall patterns that led to massive floods, pests and transformation of agricultural land into residential areas.

Domestic consumption continues to drive the local economy and private consumption continued to increase, albeit at a slower pace. As in other regions in the country, consumption spending contributed a significant share of the economy, mitigating the impact of the global financial crisis. Transfers of significant sums of money from the central government and the high amount of public savings after the reconstruction phase had helped boost local spending, despite low economic growth in the province (World Bank, 2009).

Inflationary pressures have subsided together with the global slump in demand, and inflation in Aceh has been lower than the national level since July 2008. The end of the reconstruction effort and the demand this created for goods and services, together with the slowing of Aceh's economy, have contributed to the lowest inflation rate in over four years. Before the tsunami, inflation rate in 2003 and 2004 was 3.50 percent and 6.97 percent, respectively. When the reconstruction effort started in 2005, inflation peaked at 41.11 percent. The next three years (2006-2008), inflation rate was 9.54, 11.00 and 10.27 percent. By 2009, the slowdown in reconstruction and restored supply network resulted in relatively low inflation of 3.50 percent (BPS, 2010). However, the new level of general prices remains high ever since the extreme price surge during the reconstruction period. Local people have to adjust to the new price environment including the new life style of non local people who rushed in to Aceh along with foreign donors during 2004-2006. This is partly the reason why the poverty line and the poverty incidence in Aceh continue to be high.

Indeed, the post reconstruction economy in Aceh faces a lot of challenges. The rebuilding of infrastructure and numerous sectors in Aceh led to noteworthy changes in the lifestyle of the people in the region. Income increased as a result of the reconstruction program and the inflow of funds. New cars, even luxury vehicles can be seen throughout the cities. Meanwhile, as the reconstruction effort comes to an end, many jobs will be lost. Thus, for future studies it is important to include a comprehensive social and economic analysis of the people in the villages, and to take into account parameters related to these issues in the objective function.

REFERENCES

ADB (Asian Development Bank), 2006. Progress Report Indonesia: Aceh-Nias Rehabilitation and Reconstruction. Available online: http://www.asiandevbank.org/Documents/ PAs/INO/31153-INO-PA.pdf.

Azzaino, Z., Conrad, J., Ferraro, P.J., 2002. Optimizing the Riparian Buffer: Harold Brook in the Skaneateles Lake Watershed, New York. Land Economics 78(4): 501-514.

Barakat, S., 2003. Housing Reconstruction After Conflict and Disaster. Humanitarian Practice Network, Humanitarian Policy Group, Overseas Development Institute (ODI), London.

BPS, Aceh, 2010. Inflation Rate: City of Banda Aceh, Lhokseumawe and National. Available online at <http://aceh.bps.go.id/index.php?option=com_content&view=article&id=82:laju-inflasi-kota-banda-aceh-lhokseumawe-dan-nasional&catid=29:statistik-distribusi&Itemid=38>

BRR and Partners, 2005. Aceh and Nias: One Year after the Tsunami: The Recovery Effort and Way Forward. Jakarta: BRR.

BRR and Partners, 2006. Aceh and Nias: Two Years After the Tsunami. Banda Aceh: BRR.

Byleveld, P.M., Kent, M.I., McCall, B.J., 2005. Operation Sumatra Assist: Post-tsunami Environmental and Public Health Response in Banda Aceh. ADF Health, v. 6, p. 48-53.

Duyne Barenstein, J., 2006. Housing reconstruction in post-earthquake Gujarat: A comparative analysis. Humanitarian Policy Network Paper Nr. 54. London: ODI.

Duyne Barenstein, J., Pittet, D., 2007. Post-disaster housing reconstruction. Current trends and sustainable alternatives for tsunami-affected communities in coastal Tamil Nadu. In: Point Sud. Lausanne: EPFL

ESP (Environmental Services Program), 2007. Water Quality Monitoring and Hydrochemical Loading Study Banda Aceh, Indonesia. ESP, USAID, DAI Project Number: 5300201.

Geist, E.L., 2005. Tsunamis and Earthquakes: Tsunami Generation from the 2004 Sumatra-Andaman Earthquake, USGS Western Coastal and Marine Geology. Available online at <http://walrus.wr.usgs.gov/tsunami/sumatraEQ/>

Geist, E.L., Bilek, S.L., Arcas, D., and Titov, V.V., 2006, Differences in tsunami generation between the December 26, 2004 and March 28, 2005 Sumatra earthquakes, Earth Planets Space, v. 58, p. 185-193.

ILOG, SA., 2007. CPLEX 11.0 Reference Manual.

Lay, T., H. Kanamori, C. J. Ammon, M. Nettles, S. N. Ward, R. Aster, S. L. Beck, S. L. Bilek, M. R. Brudzinski, R. Butler, H. R. DeShon, G. Ekstrom, K. Satake, and S. Sipkin., 2005. The great Sumatra-Andaman earthquake of 26 December 2004, Science, 308, 1127-1133. doi:10.1126/science.1112250.

Loomis, L., Khan, S., 2007. Journal from the Ground in Aceh Indonesia. Aceh Relief Fund. Available online at <http://www.acehrelief.org/journal/llsk/llsk050507.html>

Sphere Project, 2004. Humanitarian Charter and Minimum Standards in Disaster Response. http://www.sphereproject.org/handbook/

Srinivas, H., Nakagawa, Y., 2007. Environmental implications for disaster preparedness: Lessons Learnt from the Indian Ocean Tsunami. Journal of Environmental Management 89 (2008) 4-13. doi:10.1016/j.jenvman.2007.01.054

Steinberg, F., 2007. Housing reconstruction and rehabilitation in Aceh and Nias, Indonesia - Rebuilding lives. Habitat International, Volume 31, Issue 1, March 2007, Pages 150-166. doi:10.1016/j.habitatint.2006.11.002

Steinberg, F., Wegelin, E., Smidt, P., 2010. Conclusion and Looking Ahead: Rebuilding Lives and Homes in Aceh and Nias, Indonesia. ADB Urban Development Series.

Twigg, J., 2002. Technology, Post-Disaster Housing Reconstruction and Livelihood Security. Benfield Hazard Research Centre, Disaster Studies Working Paper 15 (January 2002/May 2006).

UNEP (United Nations Environment Programme), 2007. Environment and Reconstruction in Aceh: Two years after the tsunami. Nairobi: UNEP

Vithanage, M., Villholth, K.G., Mahatantila, K., Engesgaard, P., Jensen, K.H., 2009. Effect Of The Indian Ocean Tsunami On Groundwater Quality In Coastal Aquifers In Eastern Sri Lanka. Science of Tsunami Hazards, Vol. 28, No. 3, page 218

Walton, M., 2005. Scientists: Sumatra quake longest ever recorded. CNN. Retrieved July 2009. http://edition.cnn.com/2005/TECH/science/05/19/sumatra.quake/

World Bank, 2007. Aceh Economic Update, November 2007. Available online at <http://siteresources.worldbank.org/INTINDONESIA/Resources/226271-1176706430507/3681211-1194602678235/aeu_nov2007_en.pdf>.

World Bank, 2008. The impact of the conflict, the tsunami and reconstruction on poverty in Aceh: Aceh poverty assessment 2008. Jakarta: The World Bank.

World Bank, 2009. Aceh Economic Update, May 2009. Available online at < http://www.worldbank.org>

CHAPTER FOUR

Watershed Protection in the Skaneateles Lake

4.1 Introduction

This essay will evaluate watershed protection in the Skaneateles Lake using the clustered knapsack models developed in chapter two. Skaneateles Lake has been the primary water supply for the City of Syracuse, NY since 1894. It is one of the cleanest lakes with the best water quality in the Finger Lakes (Halfman, 2009). The lake is located approximately 20 miles southwest of the City of Syracuse. A watershed is the area of land that drains water into the lake via creeks, brooks, and drainage ways. For Skaneateles Lake, this area totals about 59 square miles. It is largely made up of agricultural and open land, but also has smaller areas of residential and commercial development.

Skaneateles Lake is approximately 15 miles long and one mile wide with a maximum depth of 300 feet. One key reason why Skaneateles Lake has high quality water is that the amount of land area that drains into the lake is relatively small (59 square miles) as compared to the surface area of the lake (14.5 square miles). The high quality of the water makes it possible to utilize the lake’s water without filtration. Skaneateles Lake is one of the few large-system surface water supplies in the country that is approved as an unfiltered water supply. The high quality of this water is due to the shape and size characteristics of the lake and watershed, the fact that sewage discharges (including from sewage treatment plants) are not allowed into surface waters in the Skaneateles lake watershed, the efforts of the City of Syracuse's watershed protection program, and the stewardship of residents and landowners of the watershed.

The City of Syracuse water system is made up of over 500 miles of pipelines to deliver water from Skaneateles Lake to the city and to distribute the water throughout the city. Water is stored in the city in the Woodland and Westcott Reservoirs on the west side of the city. Water is also stored in two standpipes and in the three tanks that comprise Morningside Reservoir.

Since 1989 the U.S. Environmental Protection Agency's (EPA) Surface Water Treatment Rule (SWTR) requires every water supplier to filter its surface water sources prior to disinfection, unless the source water meets specific water quality criteria and the supplier has developed a watershed management program. In 1991 the New York State Department of Health (DOH), which oversees the federal drinking water regulations in New York state, granted the City of Syracuse a filtration waiver by signing a Memorandum of Agreement, subject to several very strict conditions which include continuous monitoring of key water quality parameters, a back-up disinfection system, and a rigorous watershed protection program to reduce pathogen, chemical, nutrient and sediment loading into the lake. Part of the watershed management program involves the establishment of a *riparian buffer* at important areas within the watershed. A riparian buffer is a strip of land bordering a stream, lake or reservoir that intercepts and sequesters pollutant runoff (Belt et al., 1992).

One of the critical areas within the Skaneateles Lake watershed is the Town of Skaneateles, which lies at the north end of the lake. Both the Village and Town of Skaneateles prepared and updated their zoning and land use regulations, anticipating that the changes would offer improved protection of the public water supply (MacBeth, 2003). Thus, we will use the Town of Skaneateles Sub-Watershed (TSSW) as our study area. The Town of Skaneateles surrounds the intake pipes that the City of Syracuse uses to draw approximately 40.8 million gallons of water per day. It contains 1,834 parcels with 52 different land-use activities that include, among others, field crops, dairy farms, year-round residences, seasonal residences, warehouse-distribution facilities and many others.

This essay develops optimization models that decision makers can use to determine high priority parcels for inclusion in a riparian buffer. The model selects parcels based on a parcel index and a weighting of reduced pollutant loads, associated with its cluster effects.

4.2 Model and Parameters

Since September 2002, the Skaneateles Lake Watershed Land Protection Program (SLWLPP) has been actively partnering with landowners in the Skaneateles Lake watershed to protect the water quality of the lake and help maintain the lake as a source of clean drinking water. The City of Syracuse is purchasing conservation easements from willing landowners on lands that can significantly contribute to protecting water quality. Owners retain the right to farm, manage forest resources, or enjoy natural areas on their properties, while receiving fair market compensation for the value of the easement.

Azzaino et al. (2002) models the optimal riparian buffer to sequester pollutant run-off in the lake. Two models were tested, the Syracuse Scoring Equation (SSE) and the Parcel Pollutant Weighting (PPW) where each parcel is assigned biophysical attributes and equal weights of reduction in pathogens and phosphorous loading. The authors found that quantifying the desirability of outcomes based on criteria scoring and pollutant weighting approaches gave rather similar results. In related work, Ferraro (2003) investigate the targeting of conservation contracts in heterogeneous landscapes. Using biophysical and economic data from the GIS system of the City of Syracuse, New York, an environmental benefit score for conservation initiatives designed to maintain good water quality was calculated for land parcels around Lake

Skaneateles. Each parcel was scored according to its acreage, priority zoning, distance to the water intake, area of land, and length of stream footage. The work found that approaches that incorporate both biophysical and economic data are likely to generate much greater environmental benefits.

In this essay, 1,834 parcels in the Town of Skaneateles Lake Sub-Watershed (TSSW) are used as the unit of observation. We represent the Town of Skaneateles as a planar graph in a two dimension space, $G = (V, E)$. The set of parcel V corresponds to the parcels, i.e. each parcel is a member of s_v of V:

$$V = \{s_v \mid v = 1, 2, \ldots, n\}; \qquad (1)$$

where n is the number of parcels. Each parcel is evaluated based on its utility u_v and a cost c_v. Since our goal is to determine the optimal solution of the highest priority parcels for inclusion in a riparian buffer, there is a possibility that more than one parcel adjacent to others is selected, in which a cluster is formed.

In this section, we expand and modify the models used in Azzaino et al. (2002) and Ferraro (2003) to measure the potential contribution of each parcel in the system. In addition to the parcel's attributes that have been previously measured, which include parcel size (acreage), priority zone, distance to the water intake, and length of stream footage, we introduce two new attributes, slope and aspect of a parcel, and develop a parcel utility function as follows:

$$u_v = \left[w_1 \left(\frac{area}{distance} \right) + w_2.StFoot + w_3.PZone + w_4.(Slope + Aspect) \right] h_j (P_{v,j} - P_{v,j}^B) \qquad (2)$$

where u_v = Utility of parcel; w_j = proportion of utility where w_j = 0.25, j=1,2,3,4; *area* = acres of a parcel; *distance* = distance to water body from the center of parcel v;

StFoot = the length of stream footage or lake frontage in a parcel; *PZone* = priority zone; *Slope* = percent slope gradient of a parcel, *Aspect* = the direction of the slope, for example east, west, south or north; h_j = weight associated with the *j*th pollutant, where j= 1,2; $P_{v,j}$ = the loading of pollutant *j* from parcel *v* under its current land use or the potential loading if an easement is not acquired; and $P_{v,j}^{B}$= the loading of pollutant *j* from parcel *v* if the easement is acquired and parcel *v* is selected in the riparian buffer. Then $(P_{v,j} - P_{v,j}^{B}) \geq 0$ is the reduced loading of pollutant *j* from parcel *v* if the easement is purchased.

Different pollutants can contaminate surface water and groundwater supplies with varying effects. Crop nutrients from agricultural fertilizers, among others, are perhaps the most serious and widespread source of excess nitrogen and phosphorus (National Research Council 1993). Farm field runoff from crops receiving animal manure also contributes to phosphorus loading in surface waters. The amount of runoff is influenced by soil and its landscape slope (Steenhuis et al. 1981). As such we associate a weight, h_j, with the *j*th pollutant. In this exercise, h_j were set equal to one, thus $h_1 = h_2 = 1$.

The values used for the current pollutant load, $P_{v,j}$, and the reduced load, $P_{v,j}^{B}$, were based on a parcel's current land use. The Town of Skaneateles with 1834 parcels has 52 different land use classifications. The land-cover types included low-intensity residential, high-intensity residential, high-intensity commercial, pasture, row crops, mixed forest and wetlands. The New York State Office of Real Property Services has developed a simple and uniform classification system to be used in assessment administration in New York State, which corresponds closely to the City of Syracuse's land-use classifications for the 1,834 parcels in the TSSW. The nine main categories are listed in Table 4.1. The complete list of New York State property classification code can be found at

http://www.orps.state.ny.us/assessor/manuals/vol6/ref/prclas.htm.

Table 4.1. Town of Skaneateles Property Type Classification Code

Categories	Description
100	Agricultural - Property used for the production of crops or livestock
200	Residential - Property used for human habitation. Living accommodations such as hotels, motels, and apartments are in the Commercial (Cat. 400)
300	Vacant Land - Property that is not in use, is in temporary use, or lacks permanent improvement
400	Commercial - Property used for the sale of goods and/or services
500	Recreation & Entertainment - Property used by groups for recreation, amusement, or entertainment
600	Community Services - Property used for the well being of the community
700	Industrial - Property used for the production and fabrication of durable and non-durable man-made goods
800	Public Services - Property used to provide services to the general public
900	Wild, Forested, Conservation Lands & Public Parks - Reforested lands, preserves, and private hunting and fishing clubs

Source: New York State Office of Real Property Services

We applied the qualitative assessment from the New York State Office of Real Property Services to the corresponding land-use classification in the TSSW. The qualitative assessment was converted into an index number where high loading of pollutant, H=10; medium high, M(H)=8.33; medium, M=6.67; low, L=3.33; and none, N=0. The resulting indices for phosphorus and pathogens for the 52 different property classes in the TSSW are shown in Table 4.2.

Table 4.2. Land Cover and Loading of Phosphorus and Pathogens

Property Class	Land Cover		Total Phosphorus		Pathogen	
105	LC06	Row Crops	M(H)	8.33	M(H)	8.33
110	LC16	Diary Farm	H	10.00	H	10.00
112	LC16	Diary Farm	H	10.00	H	10.00
113	LC16	Diary Farm	H	10.00	H	10.00
114	LC16	Diary Farm	H	10.00	H	10.00
120	LC06	Row Crops	M(H)	8.33	M(H)	8.33
170	LC06	Row Crops	M(H)	8.33	M(H)	8.33
210	LC02	Low Intensity Residential	M	6.67	M	6.67
230	LC02	Low Intensity Residential	M	6.67	M	6.67
240	LC02	Low Intensity Residential	M	6.67	M	6.67
250	LC02	Low Intensity Residential	M	6.67	M	6.67
260	LC02	Low Intensity Residential	M	6.67	M	6.67
270	LC02	Low Intensity Residential	M	6.67	M	6.67
280	LC02	Low Intensity Residential	M	6.67	M	6.67
311	LC02	Low Intensity Residential	M	6.67	M	6.67
312	LC02	Low Intensity Residential	M	6.67	M	6.67
313	LC02	Low Intensity Residential	M	6.67	M	6.67
314	LC07	Other Grasses	L	3.33	L(M)	5.00
316	LC07	Other Grasses	L	3.33	L(M)	5.00
321	LC07	Other Grasses	L	3.33	L(M)	5.00
322	LC07	Other Grasses	L	3.33	L(M)	5.00
323	LC07	Other Grasses	L	3.33	L(M)	5.00
330	LC02	Low Intensity Residential	M	6.67	M	6.67
340	LC02	Low Intensity Residential	M	6.67	M	6.67
411	LC02	Low Intensity Residential	M	6.67	M	6.67
421	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
423	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
431	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
432	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
433	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
449	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
450	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
461	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
464	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
465	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
473	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
481	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
482	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
483	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
484	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
485	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
553	LC07	Other Grasses	L	3.33	L(M)	5.00
570	LC02	Low Intensity Residential	M	6.67	M	6.67
590	LC02	Low Intensity Residential	M	6.67	M	6.67
593	LC02	Low Intensity Residential	M	6.67	M	6.67
615	LC07	Other Grasses	L	3.33	L(M)	5.00
620	LC07	Other Grasses	L	3.33	L(M)	5.00
633	LC07	Other Grasses	L	3.33	L(M)	5.00
662	LC07	Other Grasses	L	3.33	L(M)	5.00
682	LC02	Low Intensity Residential	M	6.67	M	6.67
692	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
695	LC04	High Intensity Commercial	M(H)	8.33	M(H)	8.33
710	LC13	Barren (Quarries, mines, pits)	L	3.33	L	3.33
822	LC01	Water	L	3.33	L	3.33

Source: New York State Office of Real Property Services

The index numbers in the Total Phosphorus and Pathogen columns in Table 4.2 correspond to the current land-use loadings, $P_{v,j}$. We then make some assumptions about the potential reduction in the index numbers if a parcel were acquired for the TSSW riparian buffer. We base the percentage reductions on the discussion in Hermans (1999, p.136). When a pollutant rating is high (H), the index number is 10, we assumed that a 65% reduction was possible for parcels with a high level of phosphorus runoff. For a parcel with a high pathogen loading, a reduction of 25% would be possible under buffer status. These percentages decline to a 50% reduction for phosphorus if a parcel's current loading was low (L), and a 17% reduction for pathogens if a parcel currently had a low loading of that pollutant. The pollutant rating, index number and the percentage reduction in phosphorus and pathogens are given in Table 4.3.

Table 4.3. Percentage Reduction in Total Phosphorus and Pathogen

Pollutant Rating	Index	Percent Reduction	
		TP	Pathogen
H	10.00	65	25
M(H)	8.33	64	24.5
M	6.67	62	23
L(M)	5.00	58	21
L	3.33	50	17

Given the percentage reductions in Table 4.3, we calculated an index for pollutant loading under buffer status, corresponding to $P_{v,j}^{B}$. Index numbers by the pollutant rating is given in Table 4.4.

Table 4.4. $P_{v,j}$ and $P_{v,j}^B$ by Pollutant Rating

Pollutant Rating	Total Phosphorus		Pathogen	
	$P_{v,1}$	$P_{v,1}^B$	$P_{v,2}$	$P_{v,2}^B$
H	10.00	3.50	10.00	7.50
M(H)	8.33	3.00	8.33	6.29
M	6.67	2.53	6.67	5.13
L(M)	5.00	2.10	5.00	3.95
L	3.33	1.67	3.33	2.76

Slope and the aspect of land are important determinants for ecological site classification and natural resource management and use (Inamdar et al., 2000). For example, soil classifications often use landscape categories in the form of percent slope because it influences the pattern of native vegetation. In a similar fashion land with northern aspects produces different plant communities than land with southern aspect. The attributes aspect and slope are derived from the New York State Digital Elevation Models (DEM). A DEM contains a series of elevations ordered from south to north with the order of the columns from west to east. DEMs for this study were downloaded from the CUGIR at Cornell University website. DEMs that cover the entire Skaneateles Lake Watershed are five quadrates Q27, Q28, R27, R28, and R29. All five DEMs are merged using ESRI ArcGIS software tools. This process produced attributes Slope and Aspect for each parcel. GIS is becoming widely used to examines the contribution and complexities of the spatial dimension for analysis undertaken by environmental and resource economics (Bateman, 2002).

The *Priority Zone* is one of the criteria used by the City of Syracuse in order to determine how to effectively use its resources to protecting the watershed. The city has developed and prioritized current critical management zones identified by the process outlined in a report entitled *Land Protection Plan for the Skaneateles Lake Watershed*. The levels of Priority Zone are determined based upon both distance from the Syracuse water intakes (principally Intake #1) and distance from a watercourse or the lakeshore. The 300-foot distance from the Lake or a watercourse is based upon the definition of *environmentally sensitive area* that was drafted by the City of Syracuse. The levels of priority zone are explained in Table 4.5.

Table 4.5. Levels of Priority Zone

Level	Description
A	Within one mile of the Syracuse intakes and in a critical management zone
B	Between one and three miles of the Syracuse Intakes and in a critical management zone. Within one mile of the intakes and within 300 feet of the Lake or watercourse
C	Between three and six miles of the Syracuse intakes and in a critical management zone. Between one and three miles of the intakes and within 300 feet of a Lake or watercourse. Within one mile of the intakes and more than 300 feet from the Lake or a watercourse
D	Greater than six miles from the intakes and within 300 feet of the Lake or a watercourse. Between three and six miles of the intakes and within 300 feet of the Lake or a watercourse. Between one and three miles of the intakes and more than 300 feet from the Lake or a watercourse
E	Greater than six miles from the intakes and within 300 feet of the Lake or a watercourse. Between three and six miles of the intakes and more than 300 feet from the Lake or a watercourse
F	Greater than six miles from the intakes and more than 300 feet from the Lake or a watercourse

Source: City of Syracuse

The most popular method of retaining spatial relationships among features is to derive the adjacency information from what is known as the *topologic data model*. Topology is a mathematical concept that has its basis in the principles of feature adjacency and connectivity. Adjacency describes whether two areas are next to each other, while connectivity describes how lines are connected to each other to form a network. An adjacency attribute is included in the model to determine whether it makes any significant difference in the model performance. The adjacency attribute is an important factor for selecting riparian parcels, as it would reduce maintenance and transportation costs for the City if parcels that are adjacent to each other are selected. When a parcel is acquired, its neighbor parcels may still runoff pollutants through the acquired parcel to lake. Thus there is an essential benefit for clustering the acquired parcels to avoid such problems. For each parcel, all of these attributes were normalized to ensure that the units of measurement do not affect the parcel utility index.

The following sections describe the models used to find the optimal solution of parcels to be included in the riparian buffer. We will start by introducing the Basic Knapsack Model, which is a generalization of the knapsack problem with an underlying graph structure. We will continue with two other variants of the knapsack problem that include cluster benefit.

4.3 The Basic Knapsack Model

The Basic Knapsack Model is be formulated as follows:

$$\text{Maximize } \sum_{v} u_v x_v \tag{3}$$

$$\text{s.t. } \sum_{v} c_v x_v \leq M \tag{4}$$

$$x_v \in \{0,1\} \qquad \forall v \in V \tag{5}$$

where u_v is the utility for each parcel v which is defined in Equation (2), x_v is the binary variable associated with each parcel v (1 reflects the case where the parcel is selected, 0 is otherwise), c_v is the cost of acquiring parcel v, and M is the budget.

We refer to Equation (3) as the **Basic Knapsack Model**. We then solve the problem to optimality using IBM Ilog's Cplex solver (ILOG, 2007). The visualization of the optimal solution for the Basic Knapsack problems is shown in Figure 4.1.

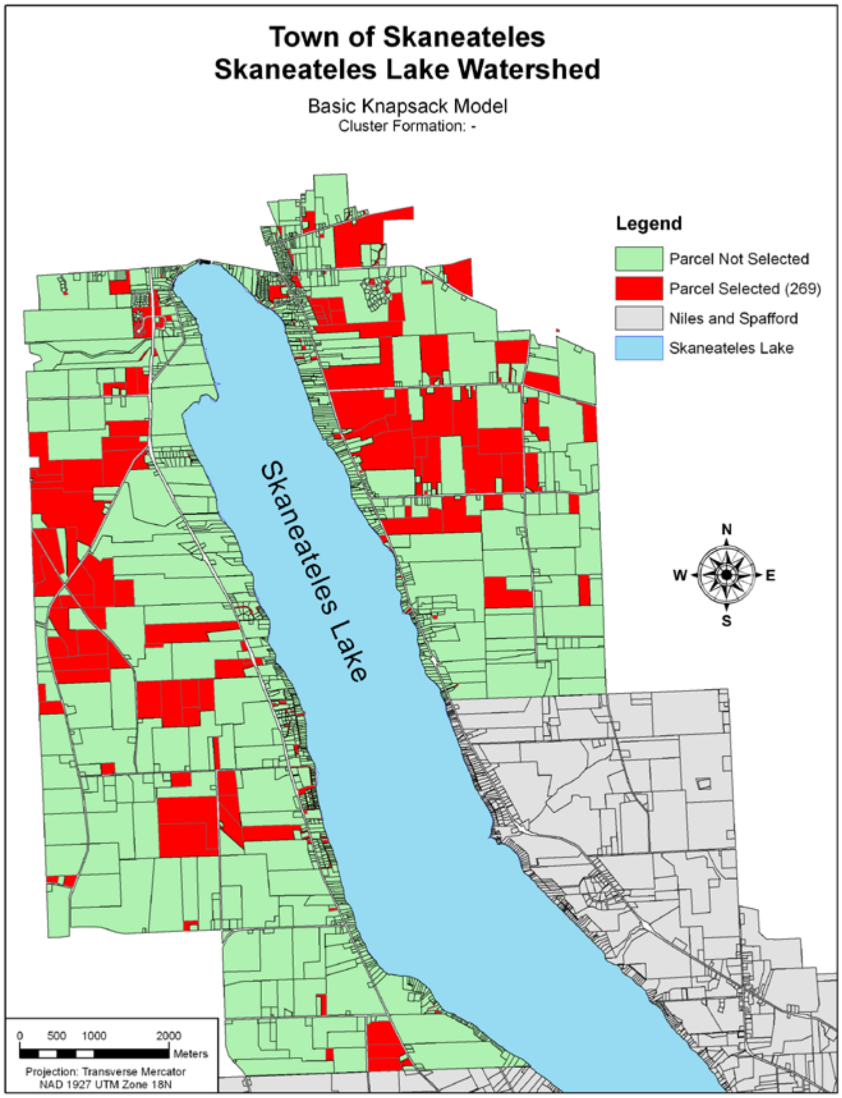


Figure 4.1. Parcels to be Acquired Based on Basic Knapsack Model

With the budget *M* of $4 million, the optimal solution for the Basic Knapsack model selected 269 parcels to be included in the buffer. Total cost to acquire the 269 parcels is $3,999,825 with total utility of 596.5. The run time needed for Cplex solver to find the optimal solution is 0.1 seconds.

4.4 The Basic Clustered Knapsack Model

As explained earlier, adjacency factor is an important factor for selecting riparian parcels, as it would reduce maintenance and transportation costs for the City of Syracuse if parcels that are adjacent to each other are selected. Furthermore, if a parcel is acquired, its neighboring parcels may still runoff pollutants through the acquired parcel to the stream or lake. Thus there is an essential benefit for clustering the acquired parcels to avoid such problem. The Basic Clustered Knapsack Model (and the next two models) is developed with this issue in mind.

In this section we introduce two types of cluster types: Neighbor and Edge. In type Neighbor, each parcel induces a cluster; the cluster is made out of the parcel itself and all the parcels adjacent to it. In type Edge, a cluster corresponds to a pair of adjacent parcels. Therefore all the clusters have size of two.

Let C be the set of m clusters in TSSW, where each cluster k is a subset of V:

$$C = \{ C_k : C_k \subseteq V \mid k = 1, 2, \ldots, m\}; \tag{6}$$

Each cluster C_k is characterized by a utility u_k. The objective is to find a set of parcels $W \subseteq V$ such that the sum of cost c_v is less or equal to the given budget M. The Basic Clustered Knapsack model can be formulated as follows:

$$\text{Maximize } \sum_k u_k y_k \tag{7}$$

$$\text{s.t. } \sum_v c_v x_v \le M \tag{8}$$

$$y_k \le x_v \qquad \forall k, v \in C_k \tag{9}$$

$$x_v \in \{0,1\} \qquad \forall v \in V \tag{10}$$

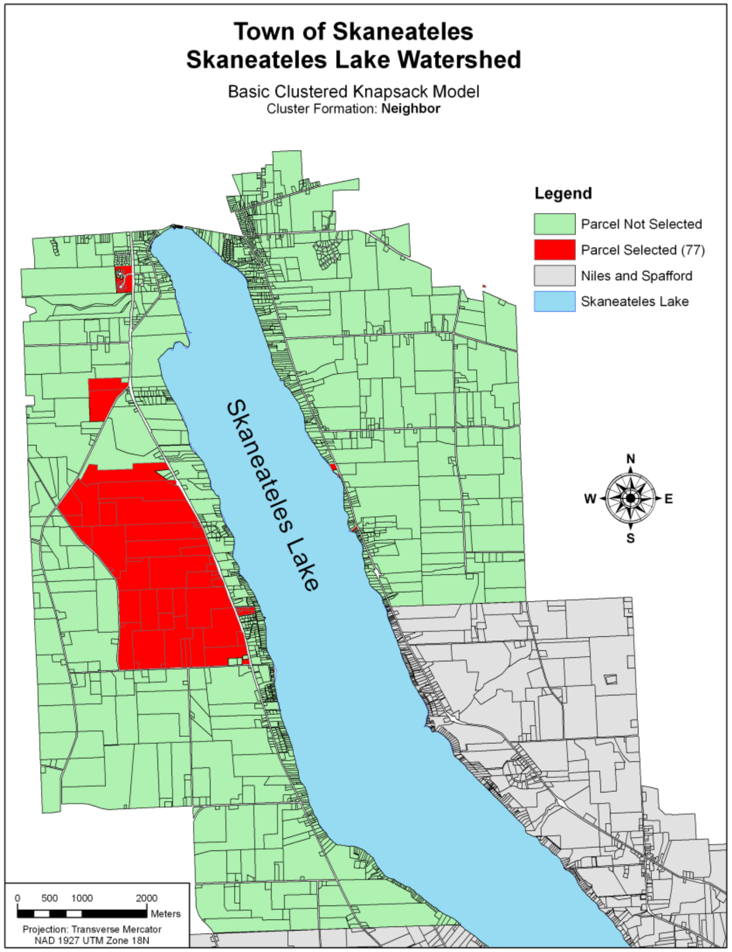
$$y_k \ge 0 \qquad \forall C_k \tag{11}$$

where u_k is the utility for each cluster k, y_k is the binary variable associated with cluster C_k (1 if all parcels in the cluster is selected, 0 otherwise), x_v is the binary variable associated with each parcel v, c_v is the cost of acquiring parcel v, and M is the budget. The cluster utility u_k is defined as follows:

$$u_k = \sum_{v \in C_k} u_v \tag{12}$$

where u_v is similarly defined as in Equation (2).

We then solve the problem to optimality using IBM Ilog's Cplex solver. The budget M is set equal to the previous model which is \$4 million. When the cluster type is Neighbor, the run time to find the optimal solution is 6.94 seconds and the number of parcels selected is only 77 parcels which total cost of \$3,999,175. This is significantly smaller than the Edge type that includes 200 parcels with total cost of slightly higher \$3,999,382 and a faster run time of 1.09 seconds. Looking at Figure 4.2, it turns out that 77 selected parcels in the Neighbor type formed one big cluster in the west part of the Skaneateles Lake, with an additional few smaller clusters, while the Edge-cluster forms somewhat medium sized clusters surrounding the lake. One might explain that when the cluster utility u_k is defined as the unit of measurement, the cost to acquire a single cluster in the Neighbor type is more expensive than type Edge, thus it forced the model to select its nearest cluster until the budget was exhausted. Visualization of optimal solution for the Basic Clustered Knapsack problems is shown in Figure 4.2.



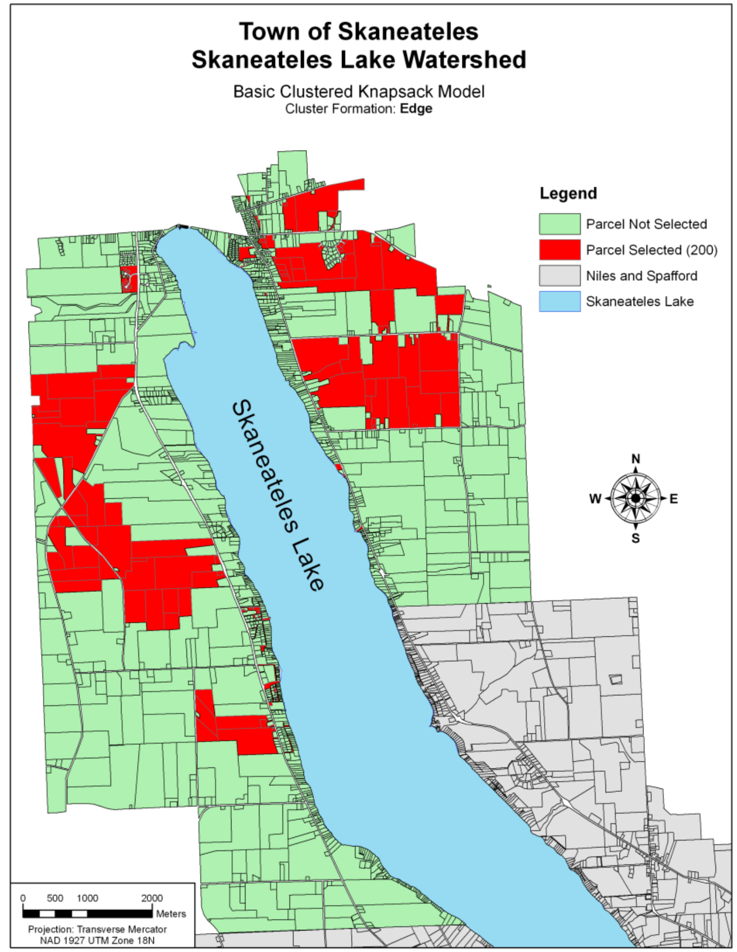


Figure 4.2. Parcels to be Acquired Based on Basic Clustered Knapsack Model: Neighbor and Edge Type

4.5 The Modified Clustered Knapsack Model

The objective function of the Clustered Knapsack Model is restated below:

$$\text{Maximize } \sum_{v} u_v x_v + \sum_{k} u_k y_k \tag{13}$$

Note that there is a possibility of double counting utility of the objective function. To avoid double counting of the utility, we replace the second term of Equation (13) with the following equation:

$$\sum_{k} \varepsilon \prod_{v \in C_k} x_v \tag{14}$$

Parameter ε is a weight to quantify the benefit of utility of producing a cluster. When all of the items in cluster k are selected, the product of x_v equals to 1, on the other hand if one or more items in cluster k are not selected, then the product of x_v equals to 0. Normally, using a product of variables in the MIP would cause the program to become nonlinear. However, we can still use the variables y_k introduced previously to indicate whether all nodes in cluster k have been selected when under the assumption that we have found an optimal solution that would push each value of y_k to the correct value. Thus Equation (14) is replaced with the following equation:

$$\sum_{k} \varepsilon y_k \tag{15}$$

The objective function in Equation (13) therefore becomes:

$$\text{Maximize } \sum_{v} u_v x_v + \sum_{k} \varepsilon y_k \tag{16}$$

$$\text{s.t.} \quad \sum_{v} c_v x_v \le M \tag{17}$$

$$y_k \le x_v;\ \forall k,\ v \in C_k \tag{18}$$

$$x_v \in \{0,1\};\ \forall v \in V \tag{19}$$

$$\mathrm{y}_k \ge 0;\ \forall C_k \tag{20}$$

Equation (16) is referred to as the **Modified Clustered Knapsack Model**. When ε is equal to zero, the objective function in Equation (16) is equal to the Basic Knapsack in Equation (3). Thus, we set the initial value of $\varepsilon = 0.01$. With all other parameters fixed, we run the model for each cluster type: Neighbor and Edge, with CPLEX by varying parameter ε from 0.01 to 1 with steps 0.01, continued with ε from 1.1 to 10 with steps 0.1, and ε from 11 to 100 with steps 1 and finally ε from 110 to 1000 with steps 10. Thus there are a total of 280 instances. We identify the value of ε that causes a change in the optimal solutions of these 280 instances. In type Neighbor, we narrowed down the solutions to 102 instances where parameter ε cause change to the optimal solutions. The running time to find the optimal solution for type Neighbor ranging between 0.56 seconds and 3.67 seconds. The first few instances of the solution were ε equals to 0.01, 0.07 0.08, 0.09, 0.22, and then the highest ε that cause change to the solutions equals to 110. This means that the optimal solutions for ε equals to 120 to 1000 are equivalent to ε=110. The two extreme optimal solutions for the Modified Clustered Knapsack model with type Neighbor are visualized in Figure 4.3. Similarly, in type Edge found 99 instances where ε cause changes in the optimal solution with running time ranging between 5.5 and 122.37 seconds. Note that when $\varepsilon \geq 110$ (110 is the higher extreme where ε cause changes) run time is significantly higher, for example when ε equals to 360 it took Cplex 8341.42 seconds (nearly 2 hours and 20 minutes) to find the optimal solution. The two extreme solutions for type Edge can be seen in Figure 4.4.

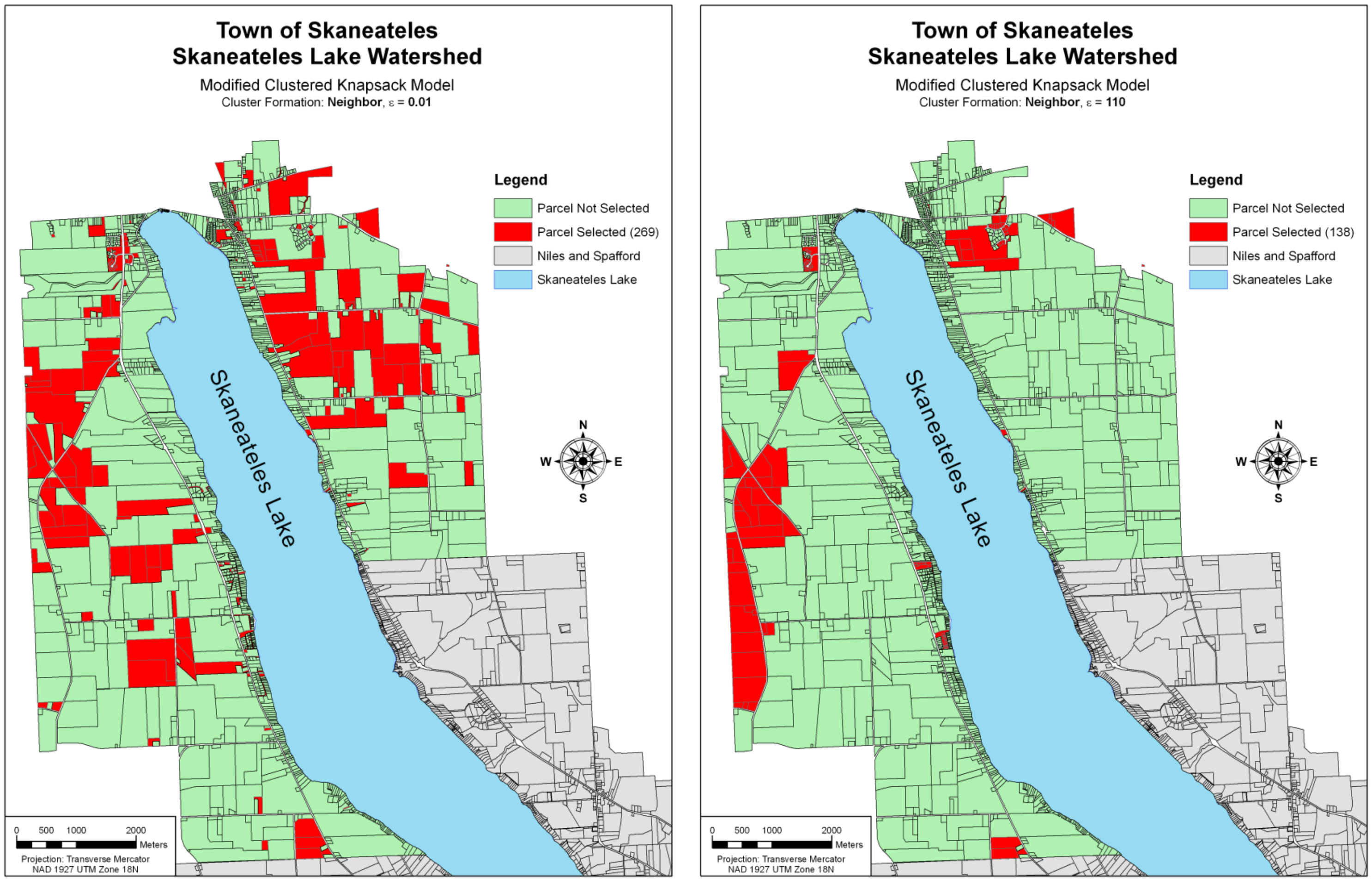
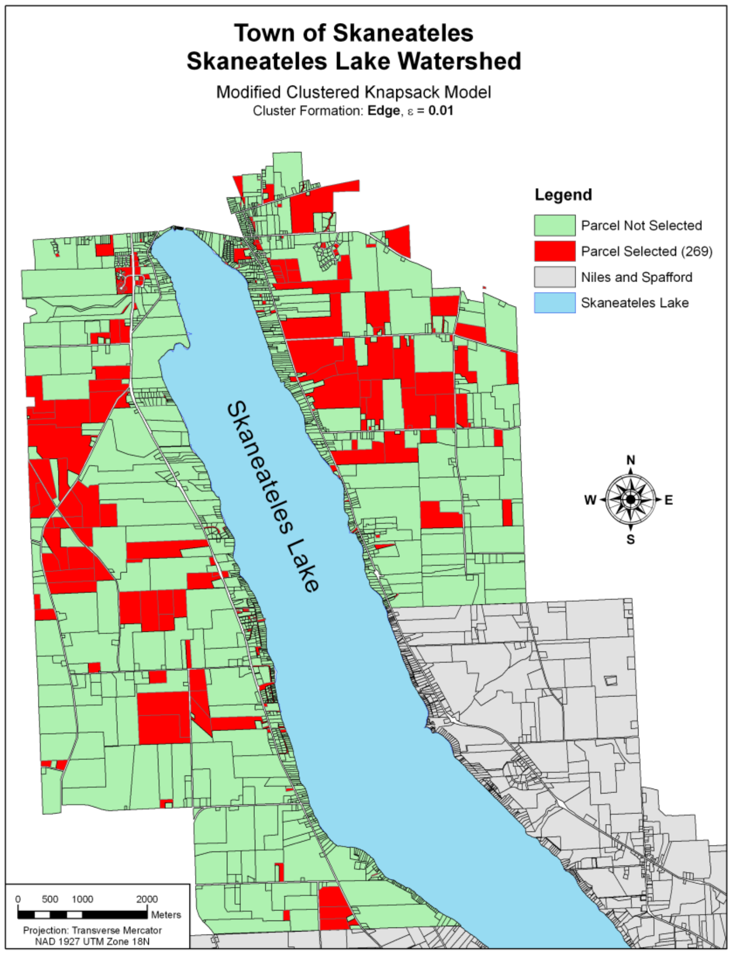


Figure 4.3. Parcels to be Acquired Based on Modified Clustered Knapsack Model: Neighbor Type



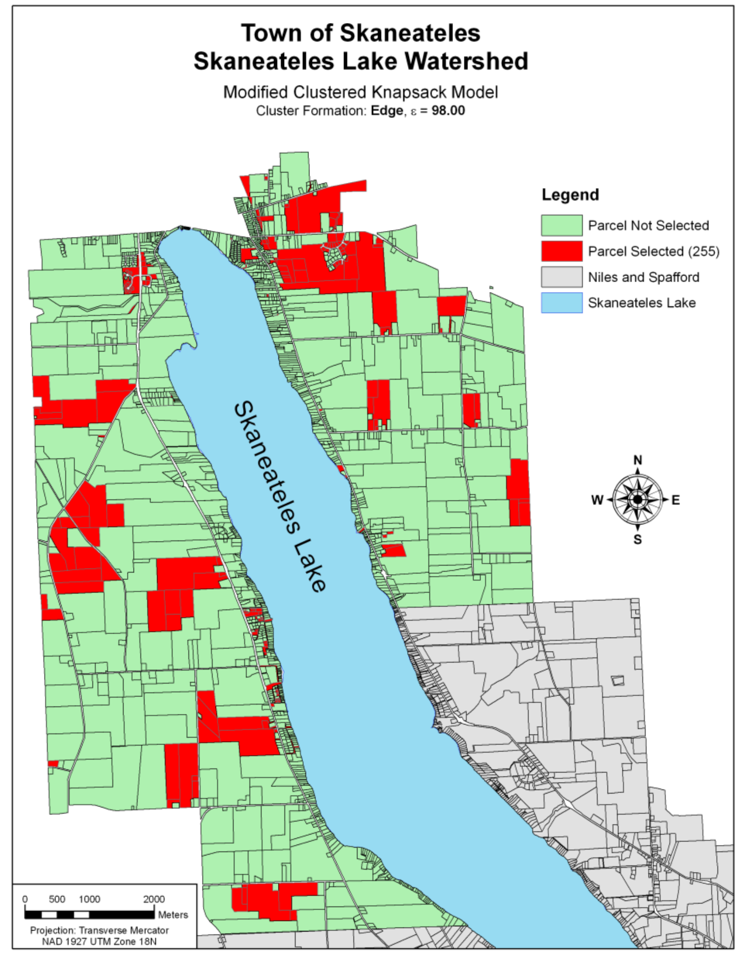


Figure 4.4. Parcels to be Acquired Based on Modified Clustered Knapsack Model: Edge Type

4.6 Conclusion and Discussion

Through the watershed protection program, agriculture continues to remain viable on properties that may have otherwise been converted to residential uses, protecting areas of prime farmland soils for future agricultural use. The placement of conservation easements on these properties protects viable farmland from extensive development that can negatively affect the Skaneateles Lake watershed while allowing farm owners to remain on their properties. The implementation of whole farm planning encourages agricultural practices that are sustainable for the land and watershed, protecting the soils and waters of the area. Because of the efforts undertaken as part of the watershed protection program, recreational uses such as fishing and swimming in the lake and hunting on the surrounding areas are protected. The result of these efforts is an enhancement of the economy, quality of life, and personal well-being of residents of the region.

A parcel utility function was developed to optimize a riparian buffer in the Skaneateles Lake watershed. Three models were constructed and applied to the Town of Skaneateles Sub-Watershed (TSSW), an area of 12,340 acres with 1834 parcels and 52 different land uses. A cluster benefit formulation was introduced in the equations. Two cluster types, Neighbor and Edge, were included in the knapsack models. Skaneateles Lake serves as the source of clean water supply for the City of Syracuse. The lake is one of the cleanest lakes with the best water quality in the Finger Lakes region such that the city is not required to build a costly filtration plant. To maintain the high quality of water, the City of Syracuse has undertaken efforts to mitigate impacts to the lake with its Skaneateles Lake Watershed Land Protection Program (SLWLPP), in addition to the Skaneateles Lake Watershed Agricultural Program (SLWAP) and ongoing education and public outreach program.

Because the City of Syracuse is not required to construct a water filtration plant for water taken from Skaneateles Lake, residents save an estimated $70 million in capital costs, and an estimated $7 million in annual operating costs. In contrast, the city spent just over $4 million to purchase conservation easements to 858 acres of land within the Skaneateles Lake Watershed (West, 2010). Of this amount, just over $3 million was fund from the city and the remaining $1 million was obtained from the US Department of Agriculture's Natural Resource Conservation Service.

We establish the optimal buffer in the TSSW under three different knapsack problems. We run the problem using CPLEX solver and find optimal buffer to the Basic Knapsack model containing 269 parcels, Basic Clustered Knapsack model with Neighbor and Edge-cluster types, 77 and 200 parcels respectively. We solve the Modified Clustered Knapsack problem with Neighbor and Edge types by varying parameter ε. The optimal buffers for type Neighbor ranging between 138 and 275 parcels. Meanwhile for type Edge we find optimal buffers ranging between 254 and 282 parcels. All problems had the same acquisition budget of $4 million. Thirty two parcels with an easement acquisition cost of $248,275, total area of 200.31 acres, and 2,300.61 total stream footage, were common to all optimal buffers. These 32 parcels are displayed in Figure 4.5 might be regarded as the *highest priority* parcels for inclusion in a riparian buffer.

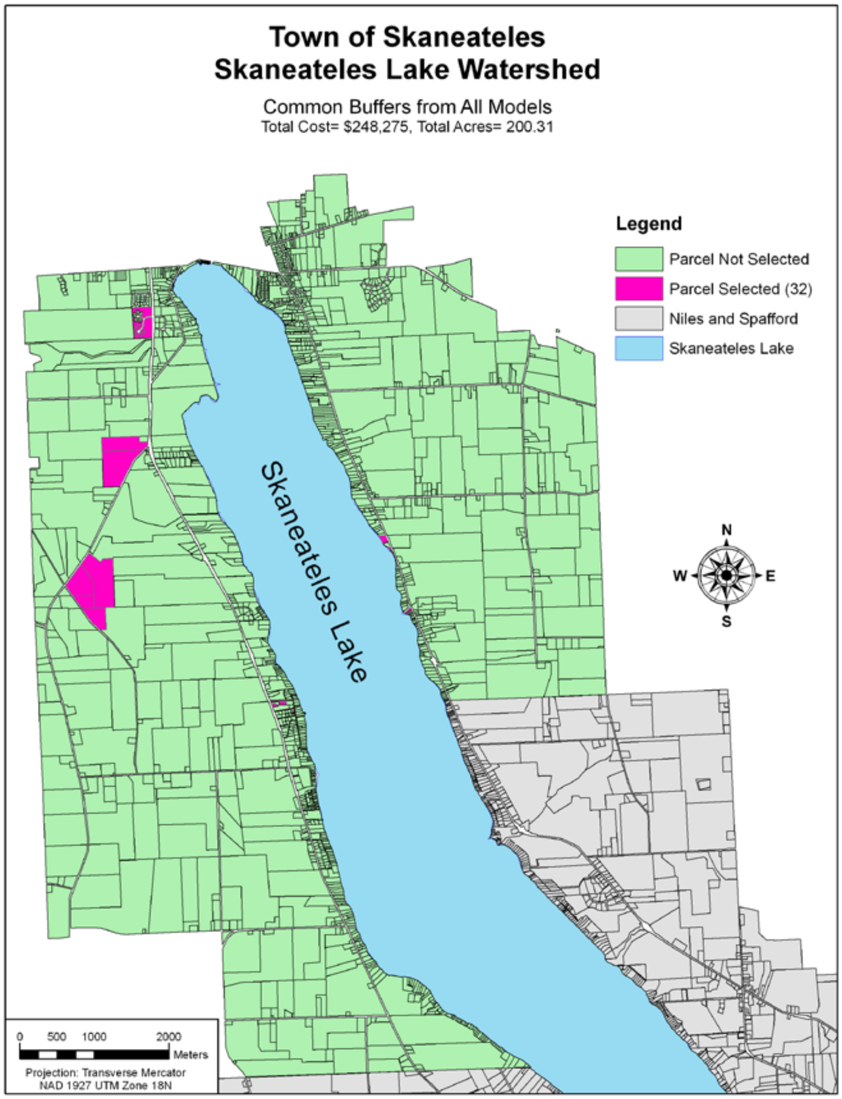


Figure 4.5. High Priority Parcels to be Acquired Based on All Models

REFERENCES

Azzaino, Z., Conrad, J., Ferraro, P.J., 2002. Optimizing the Riparian Buffer: Harold Brook in the Skaneateles Lake Watershed, New York. Land Economics 78(4): 501-514.

Bateman, I.J., A.P. Jones, A.A. Lovett, I.R. Lake and B.H. Day., 2002. Applying Geographical Information Systems (GIS) to Environmental and Resource Economics. Environmental and Resource Economics. Volume 22, Numbers 1-2, 219-269, DOI: 10.1023/A:1015575214292

Belt, G.H., O'Laughlin, J., Merrill, T., 1992. Design of Forest Riparian Buffer Strips for the Protection of Water Quality: Analysis of Scientific Literature. University of Idaho, Moscow, ID.

Ferraro, P.J., 2003. Assigning Priority to Environmental Policy Interventions in a Heterogeneous World. Journal of Policy Analysis and Management 22(1): 27-43.

Halfman, J.D., O'Neill, K., 2009. Water Quality of the Finger Lakes, New York: 2005-2008, Project Report. Finger Lakes Institute, Hobart & William Smith College. 33 pg.

Hermans, L.M., 1999. Design of Phosphorus Management Strategies for the Cannonsville Basin. New York State Water Resources Institute, Cornell University, Ithaca, New York.

ILOG, SA., 2007. CPLEX 11.0 Reference Manual.

Inamdar, S.P., Dillaha, T.A., 2000. Relationships between Drainage Area, Slope Length and Slope Gradient for Riparian Slopes in Virginia. Transactions of the ASAE, American Society of Agricultural Engineers. Vol. 43(4): 861-866.

MacBeth, L.,. 2003. New York: Skaneateles Lake Watershed - Non-point Sources Addressed in NY Watershed Management Plan. U.S. EPA #816F10033.

National Research Council, 1993. Soil and Water Quality: An Agenda for Agriculture. National Academy Press: Washington, DC.

Shepard, R., 2000. Nitrogen and phosphorus management on Wisconsin farms: Lessons learned for agricultural water quality programs. Journal of Soil and Water Conservation. 1, 63-68.

Steenhuis, T.S., Bubenzer, G.D., Converse, J.C., Walter, M.F., 1981. Winter-spread manure nitrogen loss. Transcription of the American Society of Agricultural Engineers. 24:436-449.

West, B., 2010. Project Report: Skaneateles Lake Watershed Protection Program. Behan Planning Associates, LLC.

CHAPTER FIVE

Conclusion

This dissertation is based on three related essays applying binary optimization problem, with a focus on discovering the best selection of nodes in random grids, villages in Band Aceh, and parcels in the Skaneateles Lake, that are subject to a budget constraint. The applications used in these essays are special cases of a mathematical problem called the Knapsack Problem.

The first essay investigated different mathematical formulations of the budget-constrained variant of the Clustered Knapsack Problem. In addition to the Basic Knapsack problem, we proposed three other approaches, which we refer to as Basic Clustered Knapsack, Clustered Knapsack and Modified Clustered Knapsack problem. We evaluate the performance of the models on 100 instances of a 10x10 random grids and find optimal solutions for different budget and parameter ε levels. Parameter ε is a weight to quantify the benefit of utility of producing a cluster. The results confirm that higher parameter ε causes the solution to go from mainly unclustered to highly clustered nodes, on both Neighbor and Edge types.

The second essay focused on the study of rehabilitation and reconstruction of housing in the Aceh, Indonesia, which was destroyed by the tsunami and earthquakes in 2004. Many problems arose due to lack of planning, and because not every infrastructure needed was rebuilt immediately, thousands of people remained homeless even after two years had passed. Three knapsack models from the first essay were applied to 89 villages in the City of Banda Aceh. The results discovered eight high priority villages to be rebuilt so that the residents have better access to infrastructure resources such as schools, hospitals and clean water supply. Although the post reconstruction economy in Aceh faces a lot of challenges, the rebuilding of infrastructure and numerous sectors in Aceh led to noteworthy changes in the lifestyle of the people in the region. Income increased as a result of the reconstruction program and the inflow of funds. Meanwhile, as the reconstruction effort comes to an end, many jobs will be lost. Thus, for future studies it is important to include a comprehensive social and economic analysis of the people in the villages, and to take into account parameters related to these issues in the objective function.

Finally, the third essay evaluated watershed protection in the Skaneateles Lake. The lake is the primary water supply for the city of Syracuse, NY, and is one of the cleanest lakes in the Finger Lakes. Part of the watershed management program involves the establishment of a *riparian buffer* at important areas within the watershed. Three knapsack models were used to approach the riparian buffer problem, and were applied to 1,834 parcels in the Town of Skaneateles. The results confirmed that higher multiplier ε causes the solution to become more clustered. Thirty two parcels were identified as high priority parcels for inclusion in the watershed protection program.